

Best Available Copy

SEARCH REQUEST FORM

11-38

Requestor's Name: _____ Serial Number: _____
Date: _____ Phone: _____ Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

FOR OFFICIAL USE ONLY

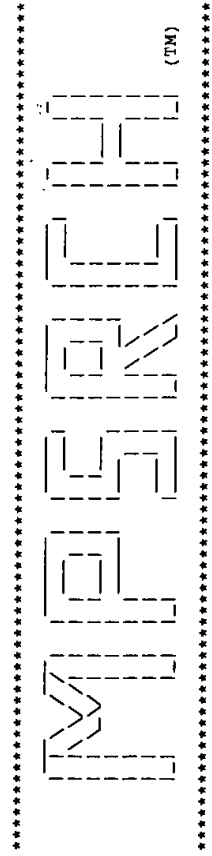
STAFF USE ONLY

Date completed: 11-10-97
Searcher: MARK
Terminal time: 6
Elapsed time: 10
CPU time: _____
Total time: 16
Number of Searches: 1
Number of Databases: 11

Search Site
☐ STIC
☒ CM-1
☐ Pre-S
Type of Search
☒ N.A. Sequence
☒ A.A. Sequence
☐ Structure
☐ Bibliographic

Vendors
☒ IG-MP RCH
☐ STN
☐ Dialog
☐ APS
☐ Geninfo
☐ SDC
☐ DARC/Questel
☐ Other

This Page Blank (uspto)



Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Thu Nov 6 10:15:18 1997; MasPar time 820.34 Seconds
Tabular output not generated. 978.029 Million cell updates/sec

Title: >US-08-842-827-1
Description: (1-1563) from US08842827.seq
Perfect Score: 1563
N.A. Sequence: 1 CCTGTGGGAGAGAGCGCGG.....CCAAAAA.....AAAAA 1563
Comp: GGACACCTCTCTCGCGGCC.....GGTTTTTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0
Searched: 707517 seqs, 256659390 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: EST-STS
1:EST1 2:EST2 3:EST3 4:EST4 5:EST5 6:EST6 7:EST7 8:EST8
9:EST9 10:EST10 11:EST11 12:EST12 13:EST13 14:EST14
15:EST15 16:EST16 17:EST17 18:EST18 19:EST19 20:EST20
21:EST21 22:EST22 23:EST23 24:EST24 25:EST25 26:EST26
27:EST27 28:EST28 29:EST29 30:EST30 31:EST31 32:EST32
33:EST33 34:EST34 35:EST35 36:EST36 37:EST37 38:EST38
39:EST39 40:EST40 41:EST41 42:EST42 43:EST43 44:EST44
45:EST45 46:EST46 47:EST47 48:EST48 49:EST49 50:EST50
51:EST51 52:EST52 53:EST53 54:EST54 55:EST55 56:EST56
57:EST57 58:EST58 59:EST59 60:EST60 61:EST61 62:EST62
63:EST63 64:EST64 65:EST65 66:EST66 67:EST67 68:EST68
69:EST69 70:EST70 71:EST71 72:EST72 73:EST73 74:EST74
75:EST75 76:EST76 77:EST77 78:EST78 79:EST79 80:EST80
81:EST81 82:EST82 83:EST83 84:EST84 85:EST85 86:EST86
87:EST87 88:EST88 89:EST89 90:EST90 91:EST91 92:EST92
93:EST93 94:EST94 95:EST95 96:EST96 97:EST97 98:EST98
99:EST99
EST-STS-TWO
100:EST100 101:EST101 102:EST102 103:EST103 104:EST104
105:EST105 106:EST106 107:EST107 108:EST108 109:EST109
110:EST110 111:EST111 112:EST112 113:EST113 114:EST114
115:EST115 116:EST116 117:EST117 118:EST118 119:EST119
120:EST120 121:EST121 122:EST122 123:EST123 124:EST124
125:EST125 126:EST126 127:EST127 128:EST128 129:EST129
130:EST130 131:EST131 132:EST132 133:EST133 134:EST134
135:EST135 136:EST136 137:EST137 138:EST138 139:EST139
140:EST140 141:EST141 142:EST142 143:EST143 144:EST144
145:EST145 146:EST146 147:EST147 148:EST148 149:EST149
150:EST150 151:EST151 152:EST152 153:EST153 154:EST154
155:EST155 156:EST156 157:EST157 158:EST158 159:EST159
160:EST160 161:EST161 162:EST162 163:EST163 164:EST164
165:EST165 166:EST166 167:EST167 168:EST168 169:EST169
170:EST170 171:EST171 172:EST172 173:EST173 174:EST174

175:EST175 176:EST176 177:EST177 178:EST178 179:EST179
180:EST180 181:EST181 182:EST182 183:EST183 184:EST184
185:EST185 186:EST186 187:EST187 188:EST188 189:EST189
190:EST190 191:EST191 192:EST192 193:EST193 194:EST194
195:EST195 196:EST196 197:EST197 198:EST198

Statistics: Mean 11.804; Variance 2.893; scale 4.080

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description	Pred. No.
C 1	528	33.8	587	75	H97570	YX03B06.s1 Homo sapie	0.00e+00
C 2	515	32.9	545	137	AA043085	ZK48B02.r1 Soares pre	0.00e+00
C 3	507	32.4	552	159	AA036943	ZK30F04.s1 Soares pre	0.00e+00
C 4	501	32.1	546	171	W67667	ZK38A11.s1 Soares pre	0.00e+00
C 5	461	29.5	465	160	AA040858	ZK48R02.s1 Soares pre	0.00e+00
C 6	448	28.7	472	114	W04968	ZK43G09.r1 Soares fet	0.00e+00
C 7	414	26.5	490	92	N31047	YX51H07.r1 Homo sapie	0.00e+00
C 8	406	26.0	446	171	W67666	ZK38A11.r1 Soares fet	0.00e+00
C 9	406	26.0	474	113	W01275	ZK40F11.r1 Soares fet	0.00e+00
C 10	405	25.9	456	169	W45106	ZC21A10.r1 Soares sen	0.00e+00
C 11	395	25.3	422	159	W45155	ZC21A10.s1 Soares sen	0.00e+00
C 12	381	24.4	456	67	H68363	YR82F11.r1 Homo sapie	0.00e+00
C 13	380	24.3	471	189	AA152123	Z1A1Q02.r1 Soares pre	0.00e+00
C 14	376	24.1	415	151	W24605	ZB63F08.r1 Soares fet	0.00e+00
C 15	377	24.1	479	74	H90961	YU86E01.r1 Homo sapie	0.00e+00
C 16	374	23.9	423	78	R97295	YK74B06.r1 Homo sapie	0.00e+00
C 17	357	22.8	452	85	H54373	YQ93G09.r1 Homo sapie	0.00e+00
C 18	357	22.8	471	85	H57343	YR05D10.r1 Homo sapie	0.00e+00
C 19	357	22.8	480	169	W45113	ZC21C09.r1 Soares sen	0.00e+00
C 20	354	22.6	404	62	H17855	YU36B12.s1 Homo sapie	0.00e+00
C 21	347	22.2	438	4	T70188	YU18C09.s1 Homo sapie	0.00e+00
C 22	347	22.2	473	85	H54296	YQ93G09.s1 Homo sapie	0.00e+00
C 23	342	21.9	356	165	C17623	Human placenta CDNA 5	0.00e+00
C 24	343	21.9	445	85	H57544	YR05D10.s1 Homo sapie	0.00e+00
C 25	334	21.4	386	5	T70311	YD26H07.s1 Homo sapie	0.00e+00
C 26	322	20.6	331	158	AA033777	ZK19G11.r1 Soares pre	0.00e+00
C 27	321	20.5	351	85	H57213	YR08D07.r1 Homo sapie	0.00e+00
C 28	319	20.4	332	158	AA033778	ZK19G11.s1 Soares pre	0.00e+00
C 29	315	20.2	408	7	T80888	YD26H07.r1 Homo sapie	0.00e+00
C 30	314	20.1	357	159	AA037575	ZK34F08.r1 Soares pre	0.00e+00
C 31	310	19.8	357	64	H26307	Y151F02.s1 Homo sapie	0.00e+00
C 32	308	19.7	386	159	AA036941	ZK30F04.r1 Soares pre	0.00e+00
C 33	304	19.4	413	67	H68364	YR82F11.s1 Homo sapie	0.00e+00
C 34	298	19.1	422	4	T69858	YU18C09.r1 Homo sapie	0.00e+00
C 35	287	18.4	304	198	H96367	YU61F11.s1 Soares pla	0.00e+00
C 36	270	17.3	279	108	HSC37E112	H. sapiens partial CD	0.00e+00
C 37	269	17.2	315	9	T90107	Y139G02.s1 Homo sapie	0.00e+00
C 38	267	17.1	287	142	N81694	L1773F Fetal heart, L	0.00e+00
C 39	257	16.4	351	36	R07363	YU96G08.s1 Homo sapie	0.00e+00
C 40	249	15.9	252	189	AA150023	Z1A1Q01.s1 Soares pre	0.00e+00
C 41	248	15.9	331	37	R09946	YF30F03.r1 Homo sapie	0.00e+00
C 42	247	15.8	325	128	W92816	ZH80G08.s1 Soares fet	0.00e+00
C 43	247	15.2	247	108	HSC36C022	H. sapiens partial CD	0.00e+00
C 44	234	15.0	511	154	AA008988	MG99E04.r1 Soares mou	0.00e+00
C 45	226	14.5	354	73	H90307	YU86E01.s1 Homo sapie	0.00e+00

ALIGNMENTS

1
LOCUS H97570 587 bp mRNA EST 12-DEC-1995
DEFINITION YX03B06.s1 Homo sapiens CDNA clone 260627 3'
ACCESSION H97570
NID g1118455
KEYWORDS EST.
SOURCE human clone-260627 primer-m13 -40 forward library-Soares melanocyte

2NDHM vector-pt7r3D (Pharmacia) with a modified polylinker host-DH10B (ampicillin resistant) Rsite1-Not I Rsite2-Eco RI Male. 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTTACATCTGAGTGGAGGCGCGCAGTCTTTTCTTTTCTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7r3 vector (Pharmacia). Library constructed by Bento Soares and M. Fatima Bonaldo. RNA from normal foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P. Albino.

ORGANISM Homo sapiens
 Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 587)
REFERENCE Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT

Contact: Wilson RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: est@watson.wustl.edu

High quality sequence stops: 351

Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES Location/Qualifiers

source 1..587

/organism="Homo sapiens"

/clone="260627"

<1..>587

BASE COUNT 173 a 117 c 120 g 174 t 3 others

Query Match 33.8%; Score 528; DB 75; Length 587;

Best Local Similarity 96.68; Pred. No. 0.00e+00;

Matches 565; Conservative 0; Mismatches 16; Indels 4; Gaps 4;

Db 1 ggctcccatcatttttaatatgtattatttttaataatcaaatcatcattcatagaagca 60

|||||

Cp 1545 GGTCTCCCATACATTTTAATATGTATTATTTAATCAAAATCATTCATAGAAGCA 1486

|||||

Db 61 tattacatcatgtttatcataagaattacattatcttttttaataaaaatgtatcacagtgg 120

|||||

Cp 1485 TATTACATACATGTTTATACATAAGCATTAACATTTTAAATAAAATGATATACAGGTGG 1426

|||||

Db 121 ggcacgttttggtagaagccttgagtttttttaattagtttagcattagataaac 180

|||||

Cp 1425 GGCACGTGTTTGGTGGAGGCTTGGAGTGTGTTTAAATGAGCTATTAGATAACC 1366

|||||

Db 181 actgagtttaagtaactatgtcacacaaagtg-gcattccaagagcattagcagca 239

|||||

Cp 1365 ACTGAGTTAAAGGTAAGTGTACACAAAGTGTGATCCCAAGAGCATATCAGCAGCA 1306

|||||

Db 240 gaagttttaagagcctgtcacacaggaaagatgcatcctcttgcttgggaaatc 299

|||||

Cp 1305 GAAGTCTTTAAAGGCTGTACACAGGAGAAAGATGATCCCTCTTGGCTTGTGGCAATC 1246

|||||

Db 300 attttcttttagaaacaggcagcttcacctgggacccctgtgctttcaaggctggg 359

|||||

Cp 1245 ATTTCCTTTAGAAAACAGGCGCAGCTTCACCTGGGACCCCTGCTGCTTTTCAAGGCTGG- 1187

|||||

Db 360 tgattgctggatgtgattccccagttgttggttttcatgcagagttgtatgagagtc 419

|||||

Cp 1186 TGATTGCTCGGATAGTAGTATCCCACTTGTGTGGTGTTCATGCAGAGTTGTATGAGAGTCC 1127

|||||

Db 420 tectcttttctcttttaaaagaagttcttctcttgaagaataatcogatacatatcacga 479

|||||

Cp 1126 TCCTCTTTTCTTTCTTTAAAGAACTTCTTCTTTTGAAGAAATCCGATACATATACAGA 1067

|||||

Db 480 actaatattgcaaccagagctccctgaatgagtcagtcacacacatcgctccagtgagg 539

|||||

Cp 1066 ACTAATATTGCAACAGAGACTCCCTGAATGAGTCCGATCACATCGCTCC-AGTGGTG 1008

|||||

Db 540 gtttataatccgnaactccgnaagagcncctccataatgggaacgg 584

|||||

Cp 1007 -TTTATAATCAGAAACTCGAAGAAAGCCACACATAAATGGATACGG 964

|||||

RESULT 2 AA043085 545 bp mRNA EST 04-SEP-1996

LOCUS ZK48b02.r1 Soares pregnant uterus NbHPU Homo sapiens CDNA clone

DEFINITION 486027 5'

ACCESSION AA043085

KEYWORDS EST.

KEYWORDS human.

ORGANISM Homo sapiens

Eukaryotae; mitochondria; eukaryotes; Metazoa; Chordata;

Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 545)

AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,

Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and

Wilson, R.

TITLE The WashU-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT Contact: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28ML3 rev2 from Amersham

High quality sequence stop: 485.

FEATURES Location/Qualifiers

source 1..545

/organism="Homo sapiens"

/note="Organ: uterus; Vector: pT7r3-Pac; Site_1: Not I;

Site_2: Eco RI; 1st strand cDNA was primed with a Not I -

oligo(dT) primer [5'

AACTGGAGAAATTCGGCGCGCTTTTCTTTTCTTTT 3'],

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT7r3 vector. Library

went through one round of normalization. Library

constructed by M. Fatima Bonaldo."

/clone="486027"

/clone.lib="Soares pregnant uterus NbHPU"

/sex="female"

/dev_stage="adult"

/lab_host="DH10B"

<1..>545

BASE COUNT 159 a 116 c 105 g 163 t 2 others

ORIGIN

Query Match 32.9%; Score 515; DB 137; Length 545;

Best Local Similarity 98.7%; Pred. No. 0.00e+00;

Matches 540; Conservative 0; Mismatches 3; Indels 4; Gaps 4;

Db 1 ccacacactgcaattggtcttggctgctatcattatgtggccttctcngntttct 60

|||||

5

vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo."

/clone="322938"
/tissue_lib="Soares senescent fibroblasts NbHSF"
/tissue_type="senescent fibroblast"
/lab_host="DH10B (ampicillin resistant)"
<1..>456

mrna 124 a 98 c 107 g 126 t 1 others
BASE COUNT 124 a 98 c 107 g 126 t 1 others
ORIGIN

Query Match 25.9%; Score 405; DB 169; Length 456;
Best Local Similarity 98.6%; Pred. No. 0.00e+00;
Matches 436; Conservative 0; Mismatches 1; Indels 5; Gaps 4;

Db 15 cttatcttcaagcaggatgaaggagactggaagactctttagcccccacactgcaa 74
QY 891 CTTATCTTCAAGCAGGATGAAGGAGACTGGGCAAGACTCTTACGCCACACTGCAA 950
Db 75 ttgtgtctgttgccttaccatttatgtgtggcctttctcgagttctgattataaacac 134
QY 951 TTTGGTCTTGTGCGGTATCTCAATTTATGTGGGCTTTTCGAGTTTCTGATTATAAACAC 1010
Db 135 cactgagagatgtgtgactgactcatttcaggagcctgtgttgcattattagttgct 194
QY 1011 CACTGGAGCGATGTGTGACTGGACTCATTCAGGAGCTCTGCTTGCATATTAGTTGCT 1070

Db 195 gtatatgtatcgattcttcaagaagaactcttttaagaagaagaagaggagac 254
QY 1071 GTATATGTATCGGATTTCTCAAGAAGAAGACTCTTTTAAAGAAGAAGAGGAGGAC 1130

Db 255 tctcatacaactctgcatgaacacacacacacactgggaactcactatccgagcaatccag 314
QY 1131 TCTCATACAACCTGTCATGAACACCAACACACACACACACACACACACACACACAC 1190

Db 315 ccttgaagaaggagcagggtgccaggtaagaactggccttgttttctaaaggaaatgat 374
QY 1191 CCTTGAAA-GGCAGCAGGTCGCCAGGTGAAGCTGGCCT-GTTTCTTAAAGGAAAAATGAT 1248

Db 375 tgcacaaggaagagatgcatcttcttctcgtggtgttgcacagccttttaagacttc 434
QY 1249 TGGCACAGCAAGAGATGATCTTTCTTCTCTCG--TGTACAAGCCTTT-AAAGACTTC 1305

Db 435 tgcctgtctatgcctcttggga 456
QY 1306 TGCTGTGATATGCCCTTTGGA 1327

RESULT 11 W45155 422 bp mrna EST 10-OCT-1996
LOCUS zc21a10.s1 Soares senescent fibroblasts NbHSF Homo sapiens CDNA
DEFINITION clone 322938 3'.
ACCESSION W45155
NID g1329246
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 Eukaryotae; mitochondria eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo. (bases 1 to 422)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 728 Std Error: 0.00

Seq primer: mob.REGA+ET

High quality sequence stop: 160.

Location/Qualifiers

1..422

/organism="Homo sapiens"

/note="vector: pT7T3D (Pharmacia) with a modified

polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco

RI; TGTACCAATCTGAAGTGGGCGCGCATTTTTTTTTTTTTTTTTT

3'), double-stranded cDNA was size selected, ligated to

Eco RI adapters (Pharmacia), digested with Not I and

cloned into the Not I and Eco RI sites of a modified pT7T3

vector (Pharmacia). Library went through one round of

normalization to a cot = 5. Library constructed by Bento

Soares and M.Fatima Bonaldo."

/clone="322938"

/clone_lib="Soares senescent fibroblasts NbHSF"

/tissue_type="senescent fibroblast"

/lab_host="DH10B (ampicillin resistant)"

complement(<1..>422)

BASE COUNT 124 a 73 c 90 g 133 t 2 others

ORIGIN

Query Match 25.3%; Score 395; DB 169; Length 422;

Best Local Similarity 98.3%; Pred. No. 0.00e+00;

Matches 415; Conservative 0; Mismatches 4; Indels 3; Gaps 3;

Db 1 ggttctcccatcattttaataatgtattattataataatacaacatcattcatagaagca 60

Cp 1545 GGTTCCTCCCATACATTTTAATATGTATTATTTAAATCAACATCATTCATAGAAGCA 1486

Db 61 tattacatacatgtttatacataagcattacatttttttaataataaaatgtatacaggtgg 120

Cp 1485 TATTACATACATGTTTATACATAAGCATTACATTTTTTTTAAATAAATGTATACAGGTGG 1426

Db 121 ggcactgttttgtaggaagcctggagtttttttaagtttagctattagataacc 180

Cp 1425 GGCATGTGTTTGGTGGAGGCTGGAGTTTTTTTAAATGAGTTAGAGCTATTAGATAACC 1366

Db 181 actgagttaaggttaactatgtcacacacaaagtgtgctcccaagagcagcagcagca 240

Cp 1365 ACTGAGTTAAAGGTAAGTAATGTACACACAAAGTGTGCATCCAGAGGCATATACACAGCA 1306

Db 241 gaagtctttaaggtctgtacaccagggaagaagatgcatcctcttgccttctgtggaatc 300

Cp 1305 GAAGTCTTTAAAGGCTTGTACACCAGGAAGAAAGATGCATCCTCTTGCCTTGTGCAATC 1246

Db 301 attttccnttagaaaaacagccagcttcaactggcagccctgctccttttcaagagctgg 360

Cp 1245 ATTTTCCCTTTAGAAAACAGGCCAGCTTCCACTGGGACCCCTCTGCTGCCCTTT-CAAGGCTGG 1187

Db 361 tggattgctcggatagtggtatcccgattcccgattgttggngtttcacgagagttgatagagt 420

Cp 1186 TG-ATGCTCGGATAGTG-ATTCCAGATGTTGTTGTTTCATGCAGAGTTGTTATGAGACT 1129

Db 421 CC 422

Cp 1128 CC 1127

RESULT 12

LOCUS H68363 456 bp mrna EST 18-OCT-1995

DEFINITION yr82f11.r1 Homo sapiens cDNA clone 211821 5'.

ACCESSION H68363

NID g1027103

KEYWORDS EST

SOURCE human clone=211821 primer=M13RP1 library=Soares fetal liver spleen

INFLS vector=pT7T3D (Pharmacia) with a modified polylinker

host=DH10B (ampicillin resistant) Rsite1=Pac I Rsite2=Eco RI Liver

IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES
Source 1..479
Location/Qualifiers
/organism="Homo sapiens"
/clone="240696"
mRNA <1..>479
BASE COUNT 131 a 99 c 102 g 142 t 5 others
ORIGIN

Query Match 24.1%; Score 377; DB 74; Length 479;
Best Local Similarity 97.9%; Pred. No. 0.00e+00;
Matches 420; Conservative 0; Mismatches 2; Indels 7; Gaps 7;
Db 1 atttgctngttgcgtatccatttatgtggccttctcgcagttctctgattataaaca 60
QY 950 ATTTGGTCTTGTGGCGGTATCCATTATGTGGGCGTTCTCGAGTTCTGATTATAACA 1009
Db 61 ccactggagcgatgtgtgactggactcattcaggagagctctggtgcaatattagttgc 120
QY 1010 CCACCTGGAGCGATGTGTGACTTGACTCATTCAGGGAGCTCTGTTGCAATATTAGTTGC 1069
Db 121 tgtatatgtatcggtattcttcaaaagaaagaaactctttaaagaaagaaagagggga 180
QY 1070 TGTATATGTATCGGATTTCTTCAAGAAAGAACTCTTTTAAAGAAAGAAAGAGGAGGA 1129
Db 181 ctctcatacaactctgcatgaacacacacaaactgggaatcactatccgagcaatcacca 240
QY 1130 CTCTCATACAACTCTGCTATGAAGAACACCACTGGGAATCAGTATCCGAGCATCACCA 1189
Db 241 gccttgaagcgagcaggggtgccaggtgaagctggtgctgttttctaaagaaagaaatgatt 300
QY 1190 GCCTTGAAGGCGACAGGGTCCCGAGGTGAAGCTGGCTGTTTCTAAAGGAAATGATT 1249
Db 301 gccacaagcaagagagatgcattcttctcctgggtgtacaagcctttaagactcttg 360
QY 1250 GCCACAAGGCAAG-AGGATGCATCTTTCTCCTGG-TGTACAAGCCTTTAAAGACTTCTG 1307
Db 361 ctgctgatgcctcttgggatggcacactttgtgtgtnacatagtttacctttnaactc 420
QY 1308 CTGCTGATATGCCCTCTTGG-ATG-CACACTTGTGTGT-ACATAGTT-ACCTTT-AACTC 1362
Db 421 agtgggttat 429
QY 1363 AGTGGTTAT 1371

Search completed: Thu Nov 6 10:31:38 1997
Job time : 980 secs.

This Page Blank (uspto)

W P S R E L A H (TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Nov 6 09:39:11 1997; MasPar time 1301.12 Seconds
Tabular output not generated. 1319.331 Million cell updates/sec

Title: >US-08-842-827-1
Description: (1-1563) from US08842827.seq
Perfect Score: 1563
N.A. Sequence: 1 CCTGTGGGAGAGCGCCGG.....CCAAAAA.....AAAAA 1563
Comp: GGACACCTCTCTCGCGCC.....GGTTTTTTTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD: Dbase 0; Query 0

Searched: 362067 seqs, 549138275 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb1-new3
1: BCT 2: FUN 3: GEN1 4: GEN2 5: HTG1 6: HTG2 7: HDM 8: INV
9: ORG 10: MAM 11: VRT 12: PLN 13: PRO 14: ROD 15: SYN 16: UNC
17: VRL

Database: genbank99
18: BCT1 19: BCT2 20: BCT3 21: BCT4 22: BCT5 23: BCT6 24: BCT7
25: BCT8 26: BCT9 27: BCT10 28: BCT11 29: GEN1 30: GEN2
31: GEN3 32: HTG1 33: HTG2 34: HTG3 35: INV1 36: INV2 37: INV3
38: INV4 39: INV5 40: INV6 41: INV7 42: INV8 43: INV9 44: INV10
45: INV11 46: MAM1 47: MAM2 48: MAM3 49: VRT1 50: VRT2 51: VRT3
52: VRT4 53: PAT1 54: PAT2 55: PAT3 56: PAT4 57: PAT5 58: PHG
59: PLN1 60: PLN2 61: PLN3 62: PLN4 63: PLN5 64: PLN6 65: PLN7
66: PLN8 67: PLN9 68: PLN10 69: PLN11 70: PRI1 71: PRI2
72: PRI3 73: PRI4 74: PRI5 75: PRI6 76: PRI7 77: PRI8 78: PRI9
79: PRI10 80: PRI11 81: PRI12 82: PRI13 83: PRI14 84: PRI15
85: ROD1 86: ROD2 87: ROD3 88: ROD4 89: ROD5 90: ROD6 91: ROD7
92: ROD8 93: STR 94: SYN 95: UNA 96: VRL1 97: VRL2 98: VRL3
99: VRL4 100: VRL5 101: VRL6 102: VRL7 103: VRL8 104: VRL9
105: VRL10
genbank-new3
106: BCT 107: GEN1 108: GEN2 109: HTG1 110: HTG2 111: INV
112: MAM 113: PHG 114: PHG 115: PLN 116: PRI1 117: PRI2
118: ROD 119: SYN 120: UNA 121: VRL
u-emb150_99
122: part1

Database: genbank-new3
106: BCT 107: GEN1 108: GEN2 109: HTG1 110: HTG2 111: INV
112: MAM 113: PHG 114: PHG 115: PLN 116: PRI1 117: PRI2
118: ROD 119: SYN 120: UNA 121: VRL
u-emb150_99
122: part1

Statistics: Mean 11.930; Variance 6.210; scale 1.921

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	577	36.9	1212	85	D84376	House mouse; Musculus	0.00e+00
2	413	26.4	1490	88	MUSPIP	Mus musculus (clone H	8.00e-297
3	130	8.3	3353	82	HUMORFA02	Human mRNA for KIAA00	1.11e-71
4	111	7.1	2206	91	RNDRI42	R.norvegicus mRNA for	2.52e-57
5	102	6.5	1444	78	HSU79294	Human clone 23748 mRN	1.29e-50
6	41	2.6	215	57	128278	Sequence 5 from paten	2.60e-08
7	34	2.2	215	57	128278	Sequence 5 from paten	3.22e-04
8	32	2.0	1727	45	SSM2RR	S.solidissima mRNA fo	4.10e-03
9	31	2.0	4601	40	DMU1584	Drosophila melanogast	1.42e-02
10	31	2.0	14626	35	EMOFLCEG	Bombyx mori fibroin 1	1.42e-02
11	31	2.0	19517	40	DMU37541	Drosophila melanogast	1.42e-02
12	30	1.9	1427	39	DDAAC3	Dictyostelium discoid	4.81e-02
13	29	1.9	1512	46	BBCA	B.taurus mRNA for cyc	1.59e-01
14	29	1.9	1837	91	RNP2XMR	R.norvegicus P2X mRN	1.59e-01
15	30	1.9	1862	64	PSMANC	Piromyces sp. mRNA fo	4.81e-02
16	29	1.9	8567	16	CAGNAT2	C.aureus GNAT2 gene	1.59e-01
17	28	1.8	363	91	RNA7CINT	R.norvegicus mRNA for	5.15e-01
18	28	1.8	417	40	DOATRICH	D.obscura A+T-rich re	5.15e-01
19	28	1.8	793	122	MDM430	Mouse Murrl mRNA, exo	5.15e-01
20	28	1.8	793	118	D85430	Mouse Murrl mRNA, exo	5.15e-01
21	28	1.8	793	14	MDM430	Mouse Murrl mRNA, exo	5.15e-01
22	28	1.8	1524	43	LPACT3	L.polyphemus mRNA for	5.15e-01
23	28	1.8	1737	39	DD19	Dictyostelium discoid	5.15e-01
24	28	1.8	2049	44	S67872	Zm-glucose-6-phosphat	5.15e-01
25	28	1.8	2630	43	PCU53325	Plasmodium chabaudi d	5.15e-01
26	28	1.8	3077	44	PFU07706	Plasmodium falciparum	5.15e-01
27	28	1.8	3774	91	RNIAC	R.norvegicus mRNA for	5.15e-01
28	28	1.8	6372	39	DDICNPA	Dictyoselium discoid	5.15e-01
29	28	1.8	10359	44	PFARPI	Plasmodium falciparum	5.15e-01
30	28	1.8	192274	110	HS435C23	Human DNA sequence **	5.15e-01
31	28	1.8	192274	5	HS435C23	Human DNA sequence **	5.15e-01
32	27	1.7	347	54	A2327	O.sativa mRNA for T23	1.62e+00
33	27	1.7	451	60	ENUS9459	Brassica napus antifur	1.62e+00
34	27	1.7	825	65	SCCOX1	Yeast cytochrome c ox	1.62e+00
35	27	1.7	1095	61	DDICSA	Dictyostelium discoid	1.62e+00
36	27	1.7	1128	43	MPUI5677	Myrmecia pilosula H18	1.62e+00
37	27	1.7	1381	39	CTHRP45MR	C.tentans mRNA for hr	1.62e+00
38	27	1.7	1702	43	PFARGPTA	Plasmodium falciparum	1.62e+00
39	27	1.7	1803	73	HSARGBPIA	H.sapiens mRNA for Ar	1.62e+00
40	27	1.7	2035	39	DDUS3884	Dictyostelium discoid	1.62e+00
41	27	1.7	2045	44	PNLATSYN	Panulirus argus gluta	1.62e+00
42	27	1.7	2798	42	HYDHP701B	Hydra magnipapillata	1.62e+00
43	27	1.7	3700	39	DDG80G	D.discoidium gp80 gen	1.62e+00
44	27	1.7	7989	44	PFU31083	Plasmodium falciparum	1.62e+00
45	27	1.7	9353	44	PFU67959	Plasmodium falciparum	1.62e+00

ALIGNMENTS

RESULT	1	D84376	1212 bp	RNA	ROD	03-OCT-1996
LOCUS						
DEFINITION						
ACCESSION						
NID						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						

REFERENCE 2 (bases 1 to 1212)
AUTHORS Kai.M., Wada, I., Imai, S., Sakane, F. and Kanoh, H.
TITLE Identification and cDNA cloning of 35-kDa phosphatidic acid
phosphatase (type 2) bound to plasma membranes. Polymerase chain
reaction amplification of mouse H202-inducible hie53 clone yielded
the cDNA encoding phosphatidic acid phosphatase
J. Biol. Chem. 271 (31), 18931-18938 (1996)
JOURNAL 96324980
MEDLINE
FEATURES
source
1..1212
/organism="Mus musculus"
/tissue_type="kidney"
287..1138
/evidence-experimental
/codon_start=1
/product="Phosphatidic acid phosphatase"
/db_xref="PID:61487873"
/translation="MFEKRLPYVALDVICVLLAGLPAILTSRHPYQRFICNDSDS
IKPIKEDTIPALGIVIPFCIIIVMSIGESLSYFNVLHNSFVGNPYATIIYKAV
GAFLEGVASQSLDIARTYIGSLRPHLEAICNPDSKINCSGDIYEDYICQNEKRV
KEGRLSFYSGHSFMYCMLFVALYQAKMGDMARLLRPMLOFLIAFSIYVGLSRV
SDYKHWSDVTVGLIQGAAMAILVALYVSDFFKTHSYKERKEEDPHITLHETASRN
YSTNEP"
BASE COUNT 277 a 316 c 313 g 306 t
ORIGIN
Query Match 36.9%; Score 577; DB 85; Length 1212;
Best Local Similarity 85.0%; Pred. No. 0.00e+00;
Matches 701; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
Db 274 ccagcgggtgacatgttcgacagacggcggtcgcgtagctggccctcgatgtgctcg 333
QY 329 CCGGGAGAGACCATGTTTGACAAGACGGCGGTGCGGTACGTGCGCTCGATGTGCTCG 388
Db 334 cgtgttgcgtggcggtgaccttttgcatttcaattcttaacagacatacccccctccagcg 393
QY 389 CGTGTGCTGGCTGGATGCTTTTGAATTTCTTACTTCAAGGCATACCCCTTCCAAGC 448
Db 394 aggaatattctgaatgacacctccatcaagctacccttcaagaagacacacatacctta 453
QY 449 AGGAGTATCTGTAAATGATGAGTCCATCAAGTACCTTACAAAGACACCATACCTTA 508
Db 454 tgccctattagtggaatagcttccattctgtattctgtattctgtattctgtattctgtatt 513
QY 509 TGGCTTATTAGTGGGAATAATCATTCATTCAGTATTCATTCATTCATTCATTCATTCAT 568
Db 514 tctgtctgttaccttaagtcttgatcgaattccctttgtggaacacccctacatacagc 573
QY 569 CCGTCTGTTTACTGTAACCTTTTGCACTCAAAATTCCTTTATCAGGAATAACTACATAGC 628
Db 574 caccatttcaaaagccgtcgagcccttttgcgagctcagctagctcagctccttgac 633
QY 629 CACTATTACAAGCCATATGGAACCTTTTATTTGGTGAGCTGCTAGTCAGTCCCTGAC 688
Db 634 tgacatcgtaagtatactataggcaggtttggcgccgaccttcttggtatctcttaaccc 693
QY 689 TGACATATGCCAAGTATTAATAGGACAGCTGCGGCTCACTTCTTGATGTTTGTGATCC 748
Db 694 agactgtcaaaaatacaactcgatgtaggtctatataggagactacatatgtcaagggaa 753
QY 749 AGATTGCTCAAAATCAACTGCAAGCAGTGGTTACATTTGAATACATATATGTCGAGGAA 808
Db 754 tgaagagaagtcgaagggcgaggtgtcttcttactcgggacactcttctctctat 813
QY 809 TGCAGAAAGAGTTAAGGAAGGCGAGTTGTCCTTCTATTACAGGCCACTCTTCGTTTCCAT 868
Db 814 gtactgcgtgtgttgcgacctttatcttcaagccaggatgaaggagactgggcaag 873
QY 869 GTACTGCATGCTGTTGTGCACTTTATCTTCAAGCCAGGATGAAGGAGACTGGGCAAG 928
Db 874 acctctacacccatcctccagtttgggtcattgtcttttccatatatgtggcccttcc 933
QY 929 ACTCTTACGCCCCCACTGCAATTTGGTCTTGTGCGGTATTCATTCATTTATGTGGCCCTTTC 988

Db 934 tcgagtgctgtactacaaacaccactgagtgacgtcacagttgactcattcaggagc 993
QY 989 TCGAGTTCTGATTTAAACACCACTGGAGCGGATGTTGACTGCATCAITCAGGAGC 1048
Db 994 tgcgtatggtatactggttgcgttgcgtatgcgtatgcgtatgcgtatgcgtatgcgtat 1053
QY 1049 TCTGGTTCGAATATTAGTTGCTGTATATGATGATGATGATGATGATGATGATGATGAT 1108
Db 1054 caaagagagaagaag 1098
QY 1109 TAAAGAAAGAAAGAGAGGAGGAGCTCTCATACAACTCTGCATGAAC 1153
RESULT 2
LOCUS MUSHIP 1490 bp mRNA ROD 11-JAN-1996
DEFINITION Mus musculus (clone HIC-53) hydrogen peroxide-inducible protein
mRNA, complete cds.
ACCESSION L43371
NID g1161099
KEYWORDS hydrogen peroxide; hydrogen peroxide-inducible protein.
SOURCE Mus musculus (clone: HIC-53) embryo calvariagag cDNA to mRNA.
ORGANISM Mus musculus
Eukaryote; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
Murinae; Mus.
REFERENCE 1 (bases 1 to 1490)
AUTHORS Egawa, K., Yoshiwara, M., Shibamura, M. and Nose, K.
TITLE Isolation of a novel ras-osteoclast gene that is induced by hydrogen
peroxide from a mouse osteoblastic cell line, MC3T3-E1
FEBS Lett. 372 (1), 74-77 (1995)
JOURNAL MEDLINE 96032549
FEATURES
Location/Qualifiers
1..1490
/organism="Mus musculus"
/clone="HIC-53"
/cell_line="MC3T3-E1"
/cell_type="osteoblast"
/dev_stage="embryo"
/tissue_type="calvariagag"
1..978
/codon_start=1
/product="hydrogen peroxide-inducible protein"
/db_xref="PID:g1161100"
/translation="ARASLELRPVGSGTGORRPIGRDEPALRQAEPAQIAASTVVK
ESRKHSGVRPIQTPGPGSRQSLAVAMGRRHGGALGICGLDRGLFLPLAP
GGDTIPVTPDFKTRLDVIVLADVICVLLAGLPAILTSRHPYQRFICNDSDSIXPY
KEDTIPYALLGGIVIPFCIIIVMSIGESLSYFNVLHNSFVGNPYATIIYKAVGAFLE
GVASQSLDIARTYIGSLRPHLEAICNPDSKINCSGDIYEDYICQNEKRVKRGRL
SFGSTLFLYVLHVAVCRTLSRRRLARLRLPMLQFLIAFSIYVGLFSV"
CDS
1..1490
3'UTR 979..1490
BASE COUNT 362 a 367 c 393 g 393 t
ORIGIN
Query Match 26.4%; Score 413; DB 88; Length 1490;
Best Local Similarity 86.8%; Pred. No. 8.00e-297;
Matches 499; Conservative 0; Mismatches 74; Indels 2; Gaps 1;
Db 321 ccagcgggtgacatgttcgacagacggcggtcgcgtagctgagctcagctcagctcagctcagct 380
QY 329 CCGGGAGAGACCATGTTTGACAAGACGGCGGTGCGGTACGTGCGCTCGATGCTCTCG 388
Db 381 cgtgttgcgtgagtgagctcgttgcgaattcttactcaaggagacatacccccctccagcg 440
QY 389 CGTGTGCTGGCTGGATGCTTTTGCATTTCTTACTTCAAGGCATACCCCTTCCAAGC 448
Db 441 aggaatattctgaatgtagctcctcctcaagccttcaagaagacacacatacctta 500
QY 449 AGGAGTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 508
Db 501 tgccttattagtggaatagctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 560

Matches 33; Conservative 75; Mismatches 83; Indels 2; Gaps 2;

Db 11 vvsrtascndkakkdgtntsdccrtwgvcdtdttrvnnndshnkysanyn9g 70
 Cp 1507 CAACATCATTCATAGAAGCATATTACATACATGTTTATACATAGCATATCATTTTT 1448
 Db 71 nnvgaakthyythnvgadskvtvdsynasgtssngtdgnr-sgadsygsstamts 129
 Cp 1447 TAATAAAAATGTATACAGGTGGGCACTGTTTGTGTGAAGGCTTGAGTTTTTTAATG 1388
 Db 130 rnrktgkannavdrnngdasvsgdkntkthakns-adgkvsgknngdrnnrygtgksn 188
 Cp 1387 AGTTAGAGCTATTAGATACCACTGAGTTAAAGGTAACATGATGTACACACAAGTGTGCA 1328
 Db 189 vsnncggnkrdv 201
 Cp 1327 TCCAAGAGGCATA 1315

RESULT 7
 LOCUS 128278 215 bp DNA PAT 30-OCT-1996
 DEFINITION Sequence 5 from patent US 5569830.
 ACCESSION 128278
 NID 91819054
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 215)
 AUTHORS Bennett, A., Labavitch, J. M., Powell, A. and Stotz, H.
 TITLE Plant inhibitors of fungal polygalacturonases and their use to control fungal disease
 JOURNAL Patent: US 5569830-A 5 29-OCT-1996;

FEATURES
 source Location/Qualifiers
 1..215
 BASE COUNT 15 a 8 c 25 g 26 t 141 others
 ORIGIN

Query Match 2.2%; Score 34; DB 57; Length 215;
 Best Local Similarity 11.5%; Pred. No. 3.22e-04;
 Matches 24; Conservative 88; Mismatches 95; Indels 2; Gaps 2;

Db 6 msssvsvrtascndkakkdntgntsswtddccrtwgvcdtdttrvnnndshnkysan 65
 Qy 39 ACCGGGAGCGCGTGCCTGGGTGAGGAGTCTG-AGGTACAGAGTGCCTGGGCTGGC 97

Db 66 ynygnnvgaakthyythnvgadskvtvdsynasgtssngtdgnrsgadysgskt 125
 Qy 98 ACACGAGCGCTCGGCACATAACCGAGTGTTCGCGGGGCTGTGAGGGAGGCGCCCGGCG 157

Db 126 amtrnrtgkannavdrnngdasvsgdkntkthaknsadgkvsgknngdrnnrygtgt 185
 Qy 158 GCCATTGCTGCGGTGGGAGCGCGCGCGG-TCTCAGCGCCCGCTCGGCTGCTCTCCTCC 216

Db 186 ksnvsnncggnkrdvssyannkcgssc 214
 Qy 217 TCCGGCTGGGAGGCGCGCTATCTCGGGGC 245

RESULT 8
 LOCUS SSM2RR 1727 bp RNA INV 18-MAY-1992
 DEFINITION S.solidissima mRNA for M2 (small) subunit of ribonucleotide reductase.
 ACCESSION X55125
 NID g10341
 KEYWORDS M2 protein; ribonucleotide reductase small subunit.
 SOURCE Atlantic surf clam.
 ORGANISM Spisula solidissima
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Mollusca; Bivalvia; Heteroconchia; Veneroidea; Mactridae; Spisula.
 REFERENCE 1 (bases 1 to 1727)

Hunt, T. and Standart, N.
 Direct Submission
 TITLE Submitted (03-NOV-1990) Hunt T., Standart N., Department of Biochemistry, Tennis Court Road, Cambridge, CB2 1QW, England
 REFERENCE 2 (bases 1 to 1727)
 AUTHORS Standart, N., Dale, M., Stewart, E. and Hunt, T.
 JOURNAL Genes Dev. 4 (1990) In press
 COMMENT Forms a complex with the product of the large M1 subunit of ribonucleotide reductase.

FEATURES
 source Location/Qualifiers
 1..1727
 /organism="Spisula solidissima"
 /dev_stage="oocyte"
 /cell_type="oocyte"
 96..1250
 /EC_number="1.17.4.1"
 /codon_start=1
 /product="M2 (small) subunit of ribonucleotide reductase"
 /db_xref="PID:g10342"
 /db_xref="SWISS-PROT:P07201"
 /translation="M2SINTRKNEISGLNGLKMKITEENKPKVILGTEINFORSTOK
 TPLKQEIKPVKKSQOVELLADNPRFVVLPIQYHDIMKMKKAEASFTAREVDLS
 KMAHWESLKKEKHFIHVLAFRAASDGIYVENLVERFSKEVQVTEACFGFQIAM
 ENHSEMYSLIDITYIKOPOERDFLNAIETPCVKKEKADAMRWINDSSSSYAEKRV
 AFAAVEGIFSGSFASIFWKKRGIMPLTFSNELSRDEGLHCDLAFCLFSLVNRV
 SOERIHOIIDEAVKIQEVLTPALPCRLIGMCDLMROYIEFVADRLLELLEKCDKLYN
 KENPDEMEHISLEGKTNFEKRVGEYQKMGVMSGNGTGDGSHATLDADF"

polyA_signal 1689..1694
 BASE COUNT 586 a 260 c 324 g 557 t
 ORIGIN

Query Match 2.0%; Score 32; DB 45; Length 1727;
 Best Local Similarity 77.6%; Pred. No. 4.10e-03;
 Matches 45; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Db 1670 tgatattaatgaatgtaataaatttaattgtgtgtaaaaaaataaaaaa 1727
 Qy 1506 TGATTTAATAATACATATTAATGTTATGGAGAACCAAAAAA 1563

RESULT 9
 LOCUS DMU11584 4601 bp DNA INV 23-JUL-1994
 DEFINITION Drosophila melanogaster Oregon-R mitochondrial A+T region.
 ACCESSION U11584
 NID g508826
 KEYWORDS mitochondrial DNA; A+T region; tandem repeats.
 SOURCE fruit fly.
 ORGANISM Mitochondrion Drosophila melanogaster
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda; Tracheata; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 4601)
 AUTHORS Lewis, D. L., Farr, C. L., Farquhar, A. L. and Kaguni, L. S.
 TITLE Sequence, Organization and Evolution of the A+T Region of Drosophila melanogaster Mitochondrial DNA
 JOURNAL Mol. Biol. Evol. 11, 523-538 (1994)
 MEDLINE 94285822
 REFERENCE 2 (bases 1 to 4601)
 AUTHORS Kaguni, L. S.
 TITLE Direct Submission
 JOURNAL Submitted (28-JUN-1994) Laurie S. Kaguni Ph.D, Dept. of Biochemistry, Michigan State University, East Lansing, MI, 48824-1318, USA

FEATURES
 source Location/Qualifiers
 1..4601
 /organism="Drosophila melanogaster"
 /mitochondrion
 /strain="Oregon-R"
 /dev_stage="embryo"
 650..1022
 /gene="mt:ori"
 /note="Description: mitochondrial origin; repeat I-A"
 /rpt_type=tandem

repeat_unit


```

intron      13506..14060
/gene="Fib-L"
/number=6
exon        14061..14499
/gene="Fib-L"
/number=7
polyA_signal 14473..14478
/gene="Fib-L"
polyA_site  14499
/gene="Fib-L"

BASE COUNT  4691 a 2726 c 2643 g 4566 t
ORIGIN
Query Match      2.0%; Score 31; DB 35; Length 14626;
Best Local Similarity 70.0%; Pred. NO. 1.42e-02;
Matches 63; Conservative 0; Mismatches 26; Indels 1; Gaps 1;

Db 3136 ataagttttatttttaataaattttataaaattattatttat-gaaattattttatata 3194
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 1536 ATACATTTTATATGTTATTTATTTTAAATCAACATCATTCATAGAAAGCATATTACATA 1477
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 3195 catattataaaatagtttttaattattt 3224
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 1476 CATGTTTATACATAGCAATTACATTTT 1447
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
LOCUS      DMU37541 19517 bp DNA INV 28-FEB-1996
DEFINITION Drosophila melanogaster mitochondrial genome, complete sequence.
ACCESSION  U37541
NID        g1166529
KEYWORDS   A-T-rich region; ATPase; circular; complete genome; cytochrome b;
cytochrome c; cytochrome c oxidase; mitochondrial DNA; NADH
dehydrogenase; ribosomal RNA; transfer RNA-Ala; transfer RNA-Arg;
transfer RNA-Asn; transfer RNA-Asp; transfer RNA-Cys; transfer
RNA-Gln; transfer RNA-Glu; transfer RNA-Gly; transfer RNA-His;
transfer RNA-Ile; transfer RNA-Leu; transfer RNA-Lys; transfer
RNA-Met; transfer RNA-Phe; transfer RNA-Pro; transfer RNA-Ser;
transfer RNA-Thr; transfer RNA-Trp; transfer RNA-Tyr; transfer
RNA-Val.
SOURCE     Drosophila melanogaster.
ORGANISM   Mitochondrion Drosophila melanogaster
Eukaryota; Insecta; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 408; 13319 to 19517)
AUTHORS    Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE      Drosophila melanogaster mitochondrial DNA: completion of the
nucleotide sequence and evolutionary comparisons
Insect Mol. Biol. 4 (4), 263-278 (1995)
MEDLINE    96423163
REFERENCE  2 (bases 404 to 5272)
AUTHORS     de Bruijn,M.H.
TITLE      Drosophila melanogaster mitochondrial DNA, a novel organization and
genetic code
Nature 304 (5923), 234-241 (1983)
MEDLINE    83245048
REFERENCE  3 (bases 5269 to 5695)
AUTHORS     Clary,D.O., Wahleithner,J.A. and Wolstenholme,D.R.
TITLE      Transfer RNA genes in Drosophila mitochondrial DNA: related 5'
flanking sequences and comparisons to mammalian mitochondrial tRNA
genes
Nucleic Acids Res. 11 (8), 2411-2425 (1983)
MEDLINE    83220794
REFERENCE  4 (bases 12511 to 12682)
AUTHORS     Clary,D.O., Goddard,J.M., Martin,S.C., Fauron,C.M. and
Wolstenholme,D.R.
TITLE      Drosophila mitochondrial DNA: a novel gene order
Nucleic Acids Res. 10 (21), 6619-6637 (1982)
MEDLINE    83090428
REFERENCE  5 (bases 804 to 1778)
AUTHORS     Satta,Y., Ishiwa,H. and Chigusa,S.I.
TITLE      Analysis of nucleotide substitutions of mitochondrial DNAs in

```

```

Drosophila melanogaster and its sibling species
Mol. Biol. Evol. 4 (6), 638-650 (1987)
MEDLINE    88174373
REFERENCE  6 (bases 441 to 2967)
AUTHORS     Satta,Y. and Takahata,N.
TITLE      Evolution of Drosophila mitochondrial DNA and the history of the
melanogaster subgroup
Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)
MEDLINE    91088557
REFERENCE  7 (bases 5268 to 13619)
AUTHORS     Careese,R.
TITLE      Drosophila melanogaster mitochondrial DNA: gene organization and
evolutionary considerations
Genetics 118 (4), 649-663 (1988)
MEDLINE    88212147
REFERENCE  8 (bases 14215 to 14512)
AUTHORS     Ballard,J.W., Olsen,G.J., Faith,D.P., Odgers,W.A., Rowell,D.M. and
Atkinson,P.W.
TITLE      Evidence from 12S ribosomal RNA sequences that onychophorans are
modified arthropods
Science 258 (5086), 1345-1348 (1992)
MEDLINE    93088057
REFERENCE  9 (bases 14917 to 19517)
AUTHORS     Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.
TITLE      Sequence, organization, and evolution of the A+T region of
Drosophila melanogaster mitochondrial DNA
Mol. Biol. Evol. 11 (3), 523-538 (1994)
MEDLINE    94285822
REFERENCE  10 (bases 1 to 19517)
AUTHORS     Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE      Direct Submission
Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department,
Michigan State University, East Lansing, MI 48824-1319, USA
FEATURES   Location/Qualifiers
            1..19517
            /organism="Drosophila melanogaster"
            /mitochondrion
            /note="derived from new and previously submitted
            sequences; sequence is a composite containing sequences
            obtained from different Drosophila melanogaster strains"
            1..65
            /gene="mt:ND6"
            /note="Description: mitochondrial NADH-ubiquinone
            oxidoreductase chain 6"
            /db_xref="FlyBase:FBgn0013685"
            /product="tRNA-Ile (GAU)"
            /complement(97..165)
            /gene="mt:ND6"
            /note="Description: mitochondrial NADH-ubiquinone
            oxidoreductase chain 6"
            /db_xref="FlyBase:FBgn0013685"
            /product="tRNA-Gln (UUG)"
            171..239
            /gene="mt:ND6"
            /note="Description: mitochondrial NADH-ubiquinone
            oxidoreductase chain 6"
            /db_xref="FlyBase:FBgn0013685"
            /product="tRNA-f-met (CAU)"
            240..1265
            /gene="mt:ND6"
            /note="Description: mitochondrial NADH-ubiquinone
            oxidoreductase chain 6"
            /db_xref="PID:g1166530"
            /transl_table=5
            /translation="MFNNSSKILFTITMITIGTLTTSNLSWLGAWMGLEINLSFIPL
            LSDNNNLMSTEASIKYFLTQVLASTVLLFSSILLMLKNNNNNEINSEFTSMINSALL
            LKGAAPFHFWMMEGLTWMMALMTWQKIAPLMISILNKIKYLLLSVILSVIL
            KAGGGLNQTSRLKLMAPFSINHLGWLMSLISEIWLILFFYSFLSFLVTFMNFIF
            KLPHLNOLFWSFVNSKILKFTLFMNFSLAGLPFLGFLPKVLYIOQLTLCNQYFMFT
            IMMSTLITLFFVYLKICYSAPFMNMFENNIMKMMNSINYNMYMTMTFFSIFGLFLI

```


RESULT	13	BBCA	1512 bp	RNA	MAM	19-JUL-1996
LOCUS		B.taurus mRNA for cyclin A.				
DEFINITION		X68321				
ACCESSION		g9				
NID		cyclin A; protein kinase activation.				
KEYWORDS		domestic cattle or domestic cow.				
SOURCE		Bos taurus				
ORGANISM		Hunt,T.				
REFERENCE		Direct Submission				
AUTHORS		Submitted (08-SEP-1992) T. Hunt, ICRF Clare Hall Laboratories,				
TITLE		South Mimms, Herts EN6 3LD, UK				
JOURNAL		2 (bases 1 to 1512)				
MEDLINE		Kobayashi,H., Stewart,E., Poon,R., Adamczewski,J.P., Gannon,J. and				
REFERENCE		Hunt,T.				
AUTHORS		Identification of the domains in cyclin A required for binding to,				
TITLE		and activation of, /p34cdc2 and p32cdk2 protein kinase subunits				
JOURNAL		Mol. Biol. Cell 3 (11), 1279-1294 (1992),				
MEDLINE		93091274				
REFERENCE		3 (bases 433 to 1221)				
AUTHORS		Brown,N.R., Noble,M.E., Endicott,J.A., Garman,E.F., Wakatsuki,S.,				
TITLE		Mitchell,E., Rasmussen,B., Hunt,T. and Johnson,L.N.				
JOURNAL		The crystal structure of cyclin A				
MEDLINE		Structure 3 (11), 1235-1247 (1995)				
COMMENT		96164440				
		In the 'Structure' publication: The 5' end of clone A3 was made by				
		PCR, and introduced a methionine in place of a valine, followed by				
		a glycine to make an NcoI site. At the 3' end, an XhoI site was				
		engineered in to allow subcloning into pET21d, which also puts 6				
		histidines at the C-terminus that were used for purification.				
FEATURES		Location/Qualifiers				
SOURCE		1..1512				
		/organism="Bos taurus"				
		/dev_stage="adult"				
CDS		/tissue_type="lymphocyte"				
		<1..1221				
		/note="subunit structure=cyclin A /p34cdc2 or cyclin A				
		/p33cdk2"				
		/codon_start=1				
		/product="Cyclin A-3"				
		/db_xref="PID:e54213"				
		/db_xref="PID:g10"				
		/db_xref="SWISS-PROT:P30274"				
		/translation="EFQEDGENVNPEKAAQAQPQPRTRAGLAVLRAGNSRGPAPQRPKY				
		RRVAPLKDDPNDIEIVVPWPKNKKQPAETIHVDEAEIEOKRTESKSESDVLAF				
		NSATLPFGPKRLADPDYMDGSFSGPHTMMSVLEDEKPVSNEVPDYHEDIHTYL				
		REMEVKCPKRGVMKKQPDITNSMRALLVDMLVEGVEEYKLQNETLHLAVNYIDRFLL				
		SMISLRGLQLVGTAAMLLASKFEETVPPVEAEFYITDDTYTKQVLRMHELVKLVLF				
		ADFLAAPTINOFLTOYFLHOOPANCVESLAMFLGELSLDLDADPYLYLPSPVTAAARF				
		HLALYTVTGQSWPESLVQKTGYILETLKPCCLDLHDYLRAPAQQQSIREKYKNSKI				
		HGVSLNNPPETLNV"				
BASE COUNT		481 a	298 c	321 g	412 t	
ORIGIN						
		Query Match	1.9%;	Score 29;	DB 46;	Length 1512;
		Best Local Similarity	72.3%;	Pred. No. 1.59e-01;		
		Matches	47;	Conservative	0;	Mismatches 18;
					Indels	0;
					Gaps	0;
Db	1448	tattaatgatgcagctgcaggataataataattgaattgttgaaaaaa	1507			
QY	1493	TATGAATCATGTTCATTAAATAATATACATATTAATAATCTATGGGAGAACCAAAAAA	1552			
Db	1508	aaaaa	1512			
QY	1553	AAAAA	1557			
RESULT	14	RNP2XMR	1837 bp	RNA	ROD	06-JAN-1995
LOCUS						

[W][A][T][E][R][M][A][N]
[S][E][Q][U][E][N][C][E]
[A][L][I][G][N][M][E][N][T]
[T][O]
[T][M]

Release 2.1D John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by Intelligenetics, Inc.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Nov 6 10:11:33 1997; MasPar time 169.52 Seconds
962.304 Million cell updates/sec

Tabular output not generated.

Title: >US-08-842-827-1
Description: (1-1563) from US08842827.seq
Perfect Score: 1563
N.A. Sequence: 1 CCTGTGGGAGAGCGCGG.....CCAAAAA.....AAAAA 1563
Comp: GGACACCTCTCTCGCGGCC.....GGTTTTTTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 142080 seqs, 52183452 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq28
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29

Statistics: Mean 9.613; Variance 6.601; scale 1.456

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES							
Result No.	Score	Query Match	Length	ID	Description	Pred. No.	
C 1	77	4.9	1047	2	Q10572	Human Natriuretic Pep	4.56e-27
C 2	74	4.7	1047	2	Q10572	Human Natriuretic Pep	2.16e-25
C 3	45	2.9	91	9	Q51746	Oligonucleotide probe	8.06e-10
C 4	45	2.9	204	1	N81164	Base substituted E.co	8.06e-10
C 5	41	2.6	91	9	Q51746	Oligonucleotide probe	8.33e-08
C 6	38	2.4	204	1	N81164	Base substituted E.co	2.50e-05
C 7	36	2.3	114	12	Q70468	Generic DNA sequence	2.32e-05
C 8	36	2.3	114	12	Q70467	Primer used in the la	2.07e-04
C 9	34	2.2	67	24	T14322	Generic DNA sequence	6.95e-05
C 10	35	2.2	114	12	Q70465	Generic DNA sequence	2.07e-04
C 11	34	2.2	114	12	Q70467	Generic DNA sequence	2.07e-04
C 12	34	2.2	114	12	Q70465	Generic DNA sequence	6.95e-05
C 13	35	2.2	114	12	Q70467	Generic DNA sequence	2.07e-04
C 14	34	2.2	114	12	Q70470	Generic DNA sequence	2.07e-04
C 15	34	2.2	114	12	Q70468	Generic DNA sequence	2.07e-04

C	16	34	2.2	114	12	Q70459	Generic DNA sequence	2.07e-04
C	17	32	2.1	198	7	Q42784	Ligand-induced gene,	6.07e-04
C	18	32	2.0	114	12	Q70466	Generic DNA sequence	1.77e-03
C	19	31	2.0	114	12	Q70470	Generic DNA sequence	5.07e-03
C	20	32	2.0	114	12	Q70466	Generic DNA sequence	1.77e-03
C	21	31	2.0	3521	7	Q41286	Clone GP3 encoding ma	5.07e-03
C	22	29	1.9	114	12	Q70472	Generic DNA sequence	4.02e-02
C	23	30	1.9	114	12	Q70472	Generic DNA sequence	1.44e-02
C	24	29	1.9	114	12	Q70473	Generic DNA sequence	4.02e-02
C	25	29	1.9	1837	23	T33852	Rat vas deferens P2x	4.02e-02
C	26	28	1.8	114	12	Q70471	Generic DNA sequence	1.11e-01
C	27	28	1.8	114	12	Q70471	Generic DNA sequence	1.11e-01
C	28	28	1.8	565	6	Q35072	HCV envelope region n	1.11e-01
C	29	28	1.8	1174	9	Q51020	Human FcER1 alpha gen	1.11e-01
C	30	26	1.7	74	21	T13613	DC43 TSAR library gen	8.06e-01
C	31	26	1.7	81	21	T13611	DC43 TSAR library gen	8.06e-01
C	32	27	1.7	347	4	Q27482	T23 cDNA.	3.02e-01
C	33	27	1.7	498	3	N50034	Sequence encoding new	3.02e-01
C	34	27	1.7	501	3	N50024	Sequence encoding new	3.02e-01
C	35	27	1.7	501	3	N50023	Sequence encoding new	3.02e-01
C	36	26	1.7	501	3	N50033	Sequence encoding new	8.06e-01
C	37	27	1.7	3520	11	Q69215	Interleukin-10 recept	3.02e-01
C	38	26	1.7	4193	14	Q82834	Human stromalin-2 DNA	8.06e-01
C	39	25	1.6	39	7	Q51787	Mixed oligonucleotide	2.12e+00
C	40	25	1.6	501	3	N50031	Sequence encoding new	2.12e+00
C	41	25	1.6	501	3	N50026	Sequence encoding new	2.12e+00
C	42	25	1.6	501	3	N50025	Sequence encoding new	2.12e+00
C	43	25	1.6	587	9	Q55685	Rabbit CAP18 gene.	2.12e+00
C	44	25	1.6	961	2	Q10791	Bovine vascular endot	2.12e+00
C	45	25	1.6	1633	17	T15608	FGF receptor ligand F	2.12e+00

ALIGNMENTS

RESULT 1

ID Q10572 standard; DNA; 1047 BP.

AC Q10572;

DT 05-APR-1991 (first entry)

DE Human Natriuretic Peptide Receptor B.

KW NPB: ANP; BNP; CNP; kidney failure; heart failure; protein kinase;

KW hyperaldosteronism; glaucoma; guanyl cyclase.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Peptide 1..22

FT Protein signal sequence

FT Protein 12

FT /label= mature NPB

FT Domain 23..455

FT /label= extracellular domain

FT /note= "binds natriuretic peptides A,B and C]"

FT Domain 456..456

FT /label= transmembrane domain

FT Domain 479..1047

FT /label= cytoplasmic domain

FT /note= "GC and protien kinase activity"

FT Modified-site 24..26

FT /label= N-glycos_site

FT Modified-site 35..37

FT /label= N-glycos_site

FT Modified-site 161..163

FT /label= N-glycos_site

FT Modified-site 195..197

FT /label= N-glycos_site

FT Modified-site 244..246

FT /label= N-glycos_site

FT Modified-site 277..279

FT /label= N-glycos_site

FT Modified-site 349..351

FT /label= N-glycos_site

FT Modified-site 600..602

FT /label= N-glycos_site

FN W09100292-A.

PD 10-JAN-1991.

PF 22-JUN-1990; U03586.
PR 23-JUN-1989; US-370673.
PA (GETH) GENENTECH INC.
PI Chang M, Goeddel D, Lowe D;
DR WPI: 91-036711/05.
DR N-PSDB; Q10324.
PT Natriuretic protein receptor B - for diagnosis and treatment of
PT kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
PS Claim 3; Fig 1; 49pp; English.
CC The sequence was derived from the DNA encoding natriuretic peptide
CC receptor B, NPRB, having guanyl cyclase (GC) activity and protein
CC kinase activity. The DNA can be inserted into expression vectors
CC for the prodn. of the protein, opt. after being mutated to produce
CC NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr=
CC 114,952). The protein (or variants) can be used in treatment of
CC natriuretic peptide disorders, and also to isolate peptides using
CC affinity chromatography. Antibodies with affinity for NPRB can
CC also be prepd.
SQ Sequence 1047 BP; 87 A; 15 C; 83 G; 51 T;
Query Match 4.9%; Score 77; DB 2; Length 1047;
Best Local Similarity 9.0%; Pred. No. 4.56e-27;
Matches 73; Conservative 213; Mismatches 517; Indels 5; Gaps 5;
Db 170 yndartddrnhvntngvnnangsnsvnhvyrnngnngnathonrangrvnycgn 229
Cp 1179 TCGGATAGTATCCCGAGTGTGGTGTTCATGCAGAGTGTATGAGATCCTCCTCT 1120
Db 230 nmhnnnnanrntngdyvnydvngsnragnratgrnwnrdtrnnanrna 289
Cp 1119 TTTCTTTTAAAGAGTCTTCTTCTTGAAGAAATCCGATACATATACAGCAACTAATA 1060
Db 290 nntvntvtrnnnnnynnnnnnrnnrnnrnnngnngnmmnnnagcnydgnnnyvnn 349
Cp 1059 TTCCACACGAGTCCCTGATGAGTCCAGTCAACACATCGCTCCAGTGGTGTATAT 1000
Db 350 ntninggrndgrnvnkmrryghvgtgnvmdkndrntdvnwamgdndsgdnnaa 409
Cp 999 CAGAACTCGAAGAGGCCACATAATGATGACGCAACAGCAACCAATTCGAGTGTG 940
Db 410 hysganknwvtrnnvkvgnnsdnncandndnscktknntstnnavangtgnnm 469
Cp 939 GGGCTA-AGAGTCTGCCAGTCTCCCTTCATCCTG-GCTTGAAGATAAAGTGCCACAAA 882
Db 470 ngvsnnnrknmmnknnasmvnrnnwonnngsnrnyhkgagrsrtnsnrgssygmt 529
Cp 881 CAGCATCGAGTACGGAACCAAGAGTGGCGCTGAATAGAGCAACCTGCTTCCTT 822
Db 530 ahgkynnnantghnkvnnvankhvkrrnnttrvnnnnkhmrdrvnnhtrnngacdn 589
Cp 821 AACTCTTTCTGCAATCCCTCGACATATGTAGTATTCAATGTACCATCGCTGCAGTTGAT 762
Db 590 nnnncvtnyrcnsndnnnd-snnndwmnrsvnnndnvkmmannhnsnshgskas 648
Cp 761 TTTTGACCAATCTGGATACAAACATCCAGAAAGTGGCGCGAGCTGCTATTGAATA 702
Db 649 ncvdsrvnnkntdygnasrntannndnnnyaknnntennnnnnnnnttgmnaadv 708
Cp 701 CTTGGCAATGCTAGTCAGGAGCTAGTACGAGCTGACCAATAAAGGTT-CCATTGG 643
Db 709 sngnnnnnnanrsgnnyngndnsknvkvngnrgnryrnandrtnnnnnnvnmnrc 768
Cp 642 CTTTGTAAATAGTGGCTATAGTATTCTCTGATAAGGAATTTGAGTCAAAAGGTTAC 593
Db 769 wandnaarndngnkngrnnrnnkngtssndnnnnnnrnnnyannnnkvnnrtaynnkr 828
Cp 582 AGTAACAGACAGGGTTCTCCAGAAATATAACGATAACTGAATGGAATGATTATTC 523
Db 829 kapannynnnhsvannkrngntvnanaandsvtnsdvngntansanstnnvntnnd 888
Cp 522 CACCTA-ATPAACGCAATAGGTATGGTCTCTTTGTAGGGTACTGTATGGACTCATCA 464
Db 889 nytendanndndvkvntngdaymvsvsgnngnrgnrgnhanhannanndavssnrrth 948

Cp 463 TTACAGAATACCTCTCGTTGGAGGGGGTATGCCCTTGAAGTAAGTAATGCCAAAGGCAAT 404
Db 949 rnhdnrrnrgvhtgnvcagvvgknmr 976
Cp 403 CCAGCCAGCAACACGACGACATCGA 376
RESULT 2
ID Q10572 standard; DNA; 1047 BP.
AC Q10572;
DT 09-APR-1991 (first entry)
DE Human Natriuretic Peptide Receptor B.
KW NPRB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
KW hyperaldosteronism; glaucoma; guanyl cyclase.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= signal sequence
FT Protein 12
FT /label= mature NPRB
FT Domain 23..455
FT /label= extracellular domain
FT /note= "binds natriuretic peptides A,B and C]"
FT Domain 456..456
FT /label= transmembrane domain
FT Domain 479..1047
FT /label= cytoplasmic domain
FT /note= "GC and protien kinase activity"
FT Modified -site 24..26
FT /label= N-glycos_site
FT Modified -site 35..37
FT /label= N-glycos_site
FT Modified -site 161..163
FT /label= N-glycos_site
FT Modified -site 195..197
FT /label= N-glycos_site
FT Modified -site 244..246
FT /label= N-glycos_site
FT Modified -site 277..279
FT /label= N-glycos_site
FT Modified -site 349..351
FT /label= N-glycos_site
FT Modified -site 600..602
FT /label= N-glycos_site
PN W09100292.A.
PD 10-JAN-1991.
PF 22-JUN-1990; U03586.
PR 23-JUN-1989; US-370673.
PA (GETH) GENENTECH INC.
PI Chang M, Goeddel D, Lowe D;
DR WPI: 91-036711/05.
DR N-PSDB; Q10324.
PT Natriuretic protein receptor B - for diagnosis and treatment of
PT kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
PS Claim 3; Fig 1; 49pp; English.
CC The sequence was derived from the DNA encoding natriuretic peptide
CC receptor B, NPRB, having guanyl cyclase (GC) activity and protein
CC kinase activity. The DNA can be inserted into expression vectors
CC for the prodn. of the protein, opt. after being mutated to produce
CC NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr=
CC 114,952). The protein (or variants) can be used in treatment of
CC natriuretic peptide disorders, and also to isolate peptides using
CC affinity chromatography. Antibodies with affinity for NPRB can
CC also be prepd.
SQ Sequence 1047 BP; 87 A; 15 C; 83 G; 51 T;
Query Match 4.7%; Score 74; DB 2; Length 1047;
Best Local Similarity 7.9%; Pred. No. 2.16e-25;
Matches 60; Conservative 215; Mismatches 479; Indels 5; Gaps 5;
Db 278 rtrnnanannrntvntvtrnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 337

CC peptide between the 2 domains. The oligonucleotides are also designed so
 CC that the expressed peptide contains 2 or 4 cysteine residues positioned
 CC in, or flanking, the unpredicted or variant residues. These residues
 CC confer some degree of conformational rigidity to the peptides. The TSARs
 CC or compans. comprising a TSAR binding domain can be used in vivo to
 CC deliver a chemically or biologically active moiety, eg. metal ion,
 CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
 CC cell. They can also replace the function of macromolecules, eg.
 CC monoclonal or polyclonal antibodies and therefore circumvent the need for
 CC complex methods of hybridoma formation or in vivo antibody production.
 CC The TSARs are easily characterised and have designed activity allowing
 CC direct and rapid detection in a screening process.
 CC Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;
 SQ

Query Match 2.2%; Score 34; DB 12; Length 114;
 Best Local Similarity 1.9%; Pred. No. 2.07e-04;
 Matches 2; Conservative 33; Mismatches 71; Indels 0; Gaps 0;

Db 3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnn 62

QY 121 GAGTGTTCGGGGGCTGTGAGGGGAGGCGCCCGGGCCATTGCTGGCGTGGAGCGC 180

Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 108

QY 181 GCGCGGTCTCAGCCGCTCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 226

RESULT 13

ID Q70465 standard; DNA; 114 BP.

AC Q70465;

DT 03-APR-1995 (first entry)

DE Generic DNA sequence to generate a random TSAR peptide library.

KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;

KW effector domain; concatenated heterofunctional protein; linker;

KW direct; rapid; detection; screening; treatment; generic; ss.

OS Synthetic.

FH Key Location/Qualifiers

FT misc.feature 55..60

FT /*tag= a

FT /*note= *this sequence represents 'z'; z can be a

FT sequence of 6, 9 or 12 nucleotides (see

FT comments)

PN W09418318-A.

PD 18-AUG-1994.

PF 01-FEB-1994; U00977.

PR 01-FEB-1993; US-013416.

PR 30-DEC-1993; US-176500.

PR 31-JAN-1994; US-189331.

PA (YUNC-) UNIV NORTH CAROLINA.

PI Fowlkes DM, Kay BK;

DR WPI; 94-279739/34.

DR P-PSDB; R65150 and R65151.

PT Identifying proteins or peptide(s) which bind a ligand - by

PT screening a recombinant vector library expressing fusion proteins

PT comprising a binding domain and an effector domain

PS Disclosure; Page 35; 255pp; English.

CC Q70465 is a generic DNA sequence used to generate random TSAR (Totally

CC Synthetic Affinity Reagents) peptides. This generic formula can also be

CC represented as follows: X(NNB)6(TGC)(NNB)12(NNB)14(TGC)(NNB)31. X

CC and Y are flanking restriction sites (X is not the same as Y) that are

CC not specified further. Other generic sequences are shown in Q70466-68.

CC Other specific peptides generated by these generic sequences are shown in

CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,

CC comprising at least two functional regions - a binding domain with

CC affinity for a ligand and a second effector peptide portion that is

CC chemically or biologically active. They may further comprise a linker

CC peptide between the 2 domains. The oligonucleotides are also designed so

CC that the expressed peptide contains 2 or 4 cysteine residues positioned

CC in, or flanking, the unpredicted or variant residues. These residues

CC confer some degree of conformational rigidity to the peptides. The TSARs

CC or compans. comprising a TSAR binding domain can be used in vivo to

CC deliver a chemically or biologically active moiety, eg. metal ion,

CC radioisotope, peptide, toxin or enzyme, to the specific target or on the

CC cell. They can also replace the function of macromolecules, eg.
 CC monoclonal or polyclonal antibodies and therefore circumvent the need
 CC for complex methods of hybridoma formation or in vivo antibody
 CC production. The TSARs are easily characterised and have designed
 CC activity allowing direct and rapid detection in a screening process.
 CC Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 2.2%; Score 35; DB 12; Length 114;

Best Local Similarity 3.7%; Pred. No. 6.95e-05;

Matches 4; Conservative 33; Mismatches 72; Indels 0; Gaps 0;

Db 3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnn 62

Cp 114 TGCCGAGCGCTGCTGTGCCAGCGCGGCGAGCTCTGTAGCCTCAGGACCTCTCTCCCGG 55

Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 111

Cp 54 CACGGCGTCCCGGTTGCTACCCCGTCCGATCCCGGCGCTCTCTCC 6

RESULT 14

ID Q70470 standard; DNA; 114 BP.

AC Q70470;

DT 10-APR-1995 (first entry)

DE Generic DNA sequence to generate a random TSAR peptide library.

KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;

KW effector domain; concatenated heterofunctional protein; linker;

KW direct; rapid; detection; screening; treatment; generic; ss.

OS Synthetic.

FH Key Location/Qualifiers

FT misc.feature 55..60

FT /*tag= a

FT /*note= *encoded by z (see comments)

PN W09418318-A.

PD 18-AUG-1994.

PF 01-FEB-1994; U00977.

PR 01-FEB-1993; US-013416.

PR 30-DEC-1993; US-176500.

PR 31-JAN-1994; US-189331.

PA (YUNC-) UNIV NORTH CAROLINA.

PI Fowlkes DM, Kay BK;

DR WPI; 94-279739/34.

DR P-PSDB; R58378.

PT Identifying proteins or peptide(s) which bind a ligand - by

PT screening a recombinant vector library expressing fusion proteins

PT comprising a binding domain and an effector domain

PS Disclosure; Page 36; 255pp; English.

CC Q70470 is a generic DNA sequence used to generate random TSAR (Totally

CC Synthetic Affinity Reagents) peptides. This generic formula can also be

CC represented as follows: X(NNB)4(CAC)(NNB)4(CAC)(NNB)8(CAC)(NNB)8

CC -(CAC)2(NNB)Y. X and Y are flanking restriction sites (X is not the same

CC as Y) that are not specified further. The peptides generated by this and

CC other generic sequences (Q70471-73) have invariant histidine residues

CC incorporated into variant sequences. TSARs are concatenated

CC heterofunctional proteins or peptides, comprising at least two functional

CC regions - a binding domain with affinity for a ligand and a second

CC effector peptide portion that is chemically or biologically active. They

CC may further comprise a linker peptide between the 2 domains. The TSARs

CC or compans. comprising a TSAR binding domain can be used in vivo to

CC deliver a chemically or biologically active moiety, eg. metal ion,

CC radioisotope, peptide, toxin or enzyme, to the specific target or on the

CC cell. They can also replace the function of macromolecules, eg.

CC monoclonal or polyclonal antibodies and therefore circumvent the need

CC for complex methods of hybridoma formation or in vivo antibody

CC production. The TSARs are easily characterised and have designed

CC activity allowing direct and rapid detection in a screening process.

CC Sequence 114 BP; 5 A; 10 C; 0 G; 0 T;

Query Match 2.2%; Score 34; DB 12; Length 114;

Best Local Similarity 9.8%; Pred. No. 2.07e-04;

Matches 11; Conservative 29; Mismatches 72; Indels 0; Gaps 0;

Db 3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnn 62

W P S R L

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_n n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Thu Nov 6 10:32:01 1997; Maspar time 427.87 Seconds
Tabular output not generated. 1003.878 Million cell updates/sec

Title: >US-08-842-827-1
Description: (1-1563) from US08842827.seq
Perfect Score: 1563
N.A. Sequence: 1 CCTGTGGGAGAGAGCGCGG.....CCAAAAA.....AAAAA 1563
Comp: GGACACCTCTCTCGCGGCC.....GGTTTTTTTTTTTTTT

Scoring table: TABLE default
Gap 6
Nmatch STD : Dbase 0; Query 0
Searched: 359085 seqs, 137405154 bases x 2
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: EST-STS-THREE
1:EST199 2:EST200 3:EST201 4:EST202 5:EST203 6:EST204
7:EST205 8:EST206 9:EST207 10:EST208 11:EST209 12:EST210
13:EST211 14:EST212 15:EST213 16:EST214 17:EST215
18:EST216 19:EST217 20:EST218 21:EST219 22:EST220
23:EST221 24:EST222 25:EST223 26:EST224 27:EST225
28:EST226 29:EST227 30:EST228 31:EST229 32:EST230
33:EST231 34:EST232 35:EST233 36:EST31 37:STS2 38:STS3
39:STS4 40:STS5 41:STS6 42:STS7 43:STS8 44:STS9 45:STS10
46:STS11 47:STS12 48:STS13
EST-STS-FOUR
49:gnEST1 50:gnEST2 51:gnEST3 52:gnEST4 53:gnEST5
54:gnEST6 55:gnEST7 56:gnEST8 57:gnEST9 58:gnEST10
59:gnEST11 60:gnEST12 61:gnEST13 62:gnEST14 63:gnEST15
64:gnEST16 65:gnEST17 66:gnEST18 67:gnEST19 68:gnEST20
69:gnEST21 70:gnEST22 71:gnEST23 72:gnEST24 73:gnEST25
74:gnEST5 75:gnEST11 76:gnEST12 77:gnEST13 78:gnEST14
79:gnEST5 80:gnEST6 81:gnEST7 82:gnEST8 83:gnEST9
84:gnEST10 85:gnEST11 86:gnEST12 87:gnEST13 88:gnEST14
89:gnEST15 90:gnEST16 91:gnEST17 92:gnEST18 93:gnEST19
94:gnEST20 95:gnEST21 96:gnEST22 97:gnEST23 98:gnEST24
99:gnEST25 100:gnEST26 101:gnEST27 102:gnEST28
103:gnEST29 104:gnEST30 105:gnEST31 106:gnEST5 107:ueEST1
108:ueEST2

Statistics: Mean 11.832; Variance 3.133; scale 3.777
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
SUMMARIES
Result Query

No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	482	30.8	514	28	AA024599	ze78b09.r1 Soares fet	0.00e+00
2	380	24.3	390	9	AA132640	z202d08.r1 Stratagene	0.00e+00
3	376	24.1	415	87	HS605336	z63f08.r1 Soares fet	0.00e+00
4	359	23.0	376	9	AA132534	z202d08.s1 Stratagene	0.00e+00
5	352	22.5	378	28	AA024600	ze78b09.s1 Soares fet	0.00e+00
6	345	22.1	377	30	AA062861	zf70e10.s1 Soares pin	0.00e+00
7	286	18.3	312	30	AA062853	zf70d09.r1 Soares pin	0.00e+00
8	259	16.6	385	33	AA107326	ml95b10.r1 Stratagene	0.00e+00
9	229	14.7	542	22	AA210789	zr90e08.r1 Soares NDH	4.23e-296
10	229	14.7	542	54	AA210789	zr90e08.r1 Soares NDH	4.23e-296
11	123	7.9	411	50	AA152779	mr78c03.r1 Stratagene	5.49e-134
12	120	7.7	241	87	HS368337	zc68b11.s1 Soares fet	1.56e-129
13	119	7.6	501	43	G23282	human STS WI-15758	4.73e-128
14	88	5.6	313	91	MM1155742	mx78g01.r1 Soares mou	6.59e-83
15	88	5.6	313	66	AA237956	mx78g01.r1 Soares mou	6.59e-83
16	76	4.9	564	1	W30942	zc64f09.r1 Soares fet	5.40e-66
17	76	4.9	564	88	HS942332	zc64f09.r1 Soares fet	5.40e-66
18	58	3.7	87	33	AA106892	ml85a07.r1 Stratagene	1.43e-41
19	53	3.4	242	27	W39815	308 Mouse VM CDNA lib	4.55e-35
20	53	3.4	337	27	W39816	309 Mouse VM CDNA lib	4.55e-35
21	53	3.4	402	27	W39811	307 Mouse VM CDNA lib	4.55e-35
22	46	2.9	320	10	AA015411	mh21g01.r1 Soares mou	2.96e-26
23	44	2.8	478	26	N75714	yv31b01.r1 Soares fet	8.23e-24
24	42	2.6	185	90	NM0272	ma69e06.r1 Soares mou	2.09e-21
25	40	2.6	184	95	NM63211	mb78e01.r1 Soares mou	4.83e-19
26	40	2.6	199	10	AA014140	mh29b02.r1 Soares mou	4.83e-19
27	40	2.6	269	95	NM5326	mb22b07.r1 Soares mou	4.83e-19
28	37	2.4	52	49	AA122946	mq69e06.r1 Stratagene	1.38e-15
29	37	2.4	52	6	AA122946	mq69e06.r1 Stratagene	1.38e-15
30	34	2.2	369	24	AA217401	mu90f02.r1 Stratagene	2.98e-12
31	34	2.2	429	56	AA144221	mr14h12.r1 Soares mou	2.98e-12
32	33	2.1	236	81	HS1146241	zp52f10.r1 Stratagene	3.59e-11
33	33	2.1	236	12	AA18434	zp52f10.r1 Stratagene	3.59e-11
34	33	2.1	301	56	AA138853	mr03a06.r1 Soares mou	3.59e-11
35	29	1.9	181	67	AA241676	MBL2SJ8B5T3 JHU96SL-B	4.97e-07
36	29	1.9	181	79	BMAA41676	MBL2SJ8B5T3 JHU96SL-B	4.97e-07
37	29	1.9	197	70	AA208192	mv91b08.r1 Guaywoodfo	4.97e-07
38	29	1.9	197	20	AA208192	mv91b08.r1 Beier mous	4.97e-07
39	29	1.9	197	103	MMAA81921	mv91b08.r1 Guaywoodfo	4.97e-07
40	29	1.9	339	1	N97773	1238C3 czapPPDd2.1, D	4.97e-07
41	30	1.9	446	12	AA181403	zp52f10.s1 Stratagene	4.90e-08
42	28	1.8	153	15	AA189532	mt90c05.r1 Soares mou	4.80e-06
43	28	1.8	244	5	AA117061	mn29d08.r1 Beddington	4.80e-06
44	28	1.8	453	105	NMW9541	mt84g07.r1 Soares mou	4.80e-06
45	28	1.8	1097	73	C22771	Dictyostellum discoid	4.80e-06

ALIGNMENTS

1
RESULT AA024599 514 bp mRNA EST 01-FEB-1997
LOCUS ze78b09.r1 Soares fetal heart NDHH19W Homo sapiens cDNA clone
DEFINITION 365081.5
ACCESSION AA024599
NID g1489504
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 514)
AUTHORS Hillier,D., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kuchaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE Washu-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT
Contact: Wilson RK
Washu-Merck EST Project
Washington University School of Medicine


```

Db 361 agagctccctgaatgagtcagtcacaca 390
      |||||
Cp 1051 AGAGCTCCTCGAATGAGTCCAGTCAACACA 1022

RESULT      3
ID HS605336 standard; RNA; EST; 415 BP.
AC W24605;
NI G1301591
DT 09-MAY-1996 (Rel. 47, Created)
DE 07-MAR-1997 (Rel. 51, Last updated, Version 2)
DE zb63f08.r1 Soares fetal lung NBHL19W Homo sapiens cDNA clone 308295
DE 5'.
DE EST.
KW Homo sapiens (human)
OS Eukaryotae; mitochondria eukaryotes; Metazoa; Chordata;
OC Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP 1-415
RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,
RA Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,
RA Parsons J., Rifkin L., Rohlfing T., Tan F., Trevaskis E.,
RA Waterston R., Williamson A., Wohlmann P., Wilson R.;
RA "WashU-Merck EST Project";
RL Unpublished.
CC Contact: Wilson RK WashU-Merck EST Project Washington University
CC School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,
CC MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:
CC est@watson.wustl.edu This clone is available royalty-free through
CC LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for
CC further information. Insert Length: 773 Std Error: 0.00 Seq primer:
CC mob.REGA+ET.
FH Key Location/Qualifiers
FH source
FH 1..415
FH /organism="Homo sapiens"
FH /note="Organ: lung; Vector: p7T73D (Pharmacia) with a
FH modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
FH strand cDNA was primed with a Not I - oligo(dT) primer
FH [5'-TGTTACCAATCTGAGTGGGAGCGCGCAATTTTCTTTTCTTTT-3'],
FH double-stranded cDNA was size selected, ligated to Eco RI
FH adapters (Pharmacia), digested with Not I and cloned into
FH the Not I and Eco RI sites of a modified p7T73 vector
FH (Pharmacia). Library went through one round of
FH normalization to a Cot = 5. Library constructed by Bento
FH Soares and M.Fatima Bonaldo. This library was constructed
FH from the same fetus as the fetal heart library, Soares
FH fetal heart NBHL19W."
FH /clone="308295"
FH /clone_lib="Soares fetal lung NBHL19W"
FH /dev_stage="19 weeks"
FH /lab_host="DH10B (ampicillin resistant)"
FH mrna
FH <1..>415
SQ Sequence 415 BP; 124 A; 87 C; 86 G; 116 T; 2 other;

Query Match 24.1%; Score 376; DB 87; Length 415;
Best Local Similarity 98.1%; Pred. No. 0.00e+00;
Matches 408; Conservative 0; Mismatches 3; Indels 5; Gaps 4;

Db 1 gagttcttgattataaacaccactggagcgatgtgtgactggactcattcaggagctc 60
Qy 991 GAGTTCTCGATTATTAACACCACTGAGCGATGTGTTGACTGGACTCATTCAGGGAGCTC 1050
Db 61 tgggtgcaatattagttctgtatgtatcggtatttttcaagaagaactcttctta 120
Qy 1051 TGGTTGCAATATTAGTTCTGTATATGTATCGGATTTCTTCAAGAAAGAACTCTTTTA 1110
Db 121 aagaaagaaagaggagactctctatacaactctgcatgaacacacacaaactggaatc 180
Qy 1111 AAGAAAGAAAGAGAGAGACTCTCTATACACTCTGCTGATGAACACCAACACTGGGAATC 1170
Db 181 actatcncagcgcaatcaccagccttgaaaggcgagcagggtgccagggtgaagctgccc 240

```

```

Qy 1171 ACTATC--C-GAGCAATCACCAGCCTTTGAAGGCAGCAGGTCGCCAGGTGAAGTCGCC 1227
      |||||
Db 241 tgtttttaaaggaatgattgccacaagn-aagaagatgcattcttctccctgggtata 299
      |||||
Qy 1228 TGTTTCTTAAAGGAATGATTGCCAAGCAAGGAGAGATGCATCTTTCTTCCTGGGTGTA 1287
      |||||
Db 300 caagcctttaaagactctgctgctgctaaatgcctcttggatgcacacctttgtgtgtaca 359
      |||||
Qy 1288 CAAGCCTTTAAAGACTTCTGCTGCTGCTGATA-TGCCCTCTTGGATGCACACTTTGTGTGTA 1346
      |||||
Db 360 tagttaccttaactcagtggtcttcttaactagctctaaactcattaaaaaaactcc 415
      |||||
Qy 1347 TAGTTACCTTTAACTCAGTGGTGTATCTAATAGCTCTAACTCATTAAAAAACTCC 1402
      |||||

RESULT      4
LOCUS AAL32534 376 bp mRNA EST 27-NOV-1996
DEFINITION z020408.s1 Stratagene colon (#937204) Homo sapiens cDNA clone
          587439 3'.
ACCESSION AAL32534
NID GI694041
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryotae; mitochondria eukaryotes; Metazoa; Chordata;
          Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 376)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
          Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
          Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E.,
          Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
          WashU-Merck EST Project
          Unpublished (1995)
TITLE Contact: Wilson RK
JOURNAL WashU-Merck EST Project
COMMENT Washington University School of Medicine
          4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
          Tel: 314 286 1800
          Fax: 314 286 1810
          Email: est@watson.wustl.edu
          This clone is available royalty-free through LLNL; contact the
          IMAGE Consortium (info@image.llnl.gov) for further information.
          Seq primer: -40M13 fwd from Amersham
          High quality sequence stop: 285.
FEATURES
          Location/Qualifiers
          source
          1..376
          /organism="Homo sapiens"
          /note="Organ: colon; Vector: pBluescript SK-; Site_1:
          EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
          Oligo dT. T-84 colonic epithelial cell line. Average
          insert size: 1.0 kb; Uni-ZAP XR vector; -5' adaptor
          sequence: 5' GAATTCGGCAGGAG 3' -3' adaptor sequence: 5'
          CTCGAGTCTTTTCTTTTCTTTTCTTTT 3'"
          /clone="587439"
          /clone_lib="Stratagene colon (#937204)"
          /lab_host="SOAR cells (kanamycin resistant)"
          complement(<1..>376)
          BASE COUNT 107 a 78 c 83 g 108 t
          mrna
          ORIGIN
          Query Match 23.0%; Score 359; DB 9; Length 376;
          Best Local Similarity 98.9%; Pred. No. 0.00e+00;
          Matches 373; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

Db 1 gagcgatgtgttgactgactcattcaggagctctgtgtgcaatattagttgctgtata 60
Qy 1016 GAGCGATGTGTGACTGGACTCATTCAGGGAGCTCTGGTTCATATATTAGTTGCTGTATA 1075
Db 61 tgtatcgatttcttcaagaagaactcttcttaagaagaagaagaggagactctca 120
Qy 1076 TGTATCGATTCTTCTCAAGAAAGAACTCTTTTAAAGAAAGAAAGAGGAGGACTCTCA 1135

```

```
Db 121 tacaactctgcatgaacacacaaactgggaatcactatccgagcaatcaccagccttg 180
|||||
QY 1136 TACAACCTCTGCATGAACACCAACCACTGGGATCACTATCCGAGCAATCACCAGCCTTG 1195
|||||
Db 181 aaaggcagcagggtgccagcagtggaagcctggttttctaaaggaaaatgattgccaca 240
|||||
QY 1196 AAAGGCAGCAGGGTGCCAGCGTGAAGCTGGCTGTTTCTAAAGGAAATGATTGCCACA 1255
|||||
Db 241 agg-aagaggatcactcttctccctgggtgtacaaagcctttaagactctctgctgtct 299
|||||
QY 1256 AGCAGAGGATGCATCTTCTTCCTGGGTGTACAAAGCTTTAAAGACTTCTGCTGTGAT 1315
|||||
Db 300 atgccccttgatgcacactttgtgtgtacatagttaccctttaactcagtggttatcta 359
|||||
QY 1316 ATGCCCTCTGGATGCACACTTCTGTGT-ACATAGTTACCTTTAACTCAGTGGTTATCTA 1374
|||||
Db 360 atagctctaaactcctt 376
|||||
QY 1375 ATAGCTCTAAACTCAAT 1391
|||||

RESULT 5
LOCUS AA024600 378 bp mRNA EST 01-FEB-1997
DEFINITION ze78b09.s1 Soares fetal heart NBHH19W Homo sapiens cDNA clone
365081 3'.
ACCESSION AA024600
NID G1489505
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 378)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le.M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
WashU-Merck EST Project
Unpublished (1995)

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 600 Std Error: 0.00
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 341.
Location/Qualifiers
1..378
/organism="Homo sapiens"
/notes="Organ: heart; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTCACCAATCTGAATGAGCGCCGATCTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBHL19W."
/clone="365081"
/clone_lib="Soares fetal heart NBHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"

FEATURES
source
```

```
mrna complement(<1..>378)
BASE COUNT 114 a 68 c 75 g 121 t
ORIGIN

Query Match 22.5%; Score 352; DB 28; Length 378;
Best Local Similarity 99.2%; Pred. NO. 0.00e+00;
Matches 365; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

Db 1 ttgggtctccatcacattttaataatgtattatatatttaataatcaaacatcatctcatagaaaag 60
|||||
Cp 1547 TTGGTTCTCCCATACATTTTAATATGTATATATTAAATCAACATCATTCATAGAAAG 1488
|||||
Db 61 catattacatacatgtttatatacataagcattacatttttttaataaaaaatgatacaggt 120
|||||
Cp 1487 CATATTACATACATGTTTATACATAAAGCATTTACATTTTAAATAAATGTTATACAGGT 1428
|||||
Db 121 ggggcaactgttttgggtggaagccttggaggttttttaattggaggttttagagctattagata 180
|||||
Cp 1427 GGGGCACACTGTTTGTGGTGAAGGCTTGGAGCTTTTAAATG-AGTTTAGAGCTATTAGATA 1369
|||||
Db 181 accactgagttaaaggtaactatgtacacacaaagtgtgcatcccaaggagagcagatagcgc 240
|||||
Cp 1368 ACCACTGAGTTAAAGGTAACATATGTACACACAAAGTGTGCATCCCAAGAGGCATATCA-GC 1310
|||||
Db 241 agcagaagtctttaaaggccttgcacacaggaagaaagatgcattctcttgccttgggc 300
|||||
Cp 1309 AGCAGAAAGTCCTTTAAAGGCTTGTACACAGGAAAGAAATGATGCTCTTGGCTTGTGGC 1250
|||||
Db 301 aatcatttcccttagaaaacagggccagcttcacctcctggcaccctgcctgccttcaaggc 360
|||||
Cp 1249 AATCATTTTCTTTAGAAAAACAGGCAGCTTCACCTGGGCACCTGCTGCCCTTCAAGGC 1190
|||||
Db 361 tgggtgatt 368
|||||
Cp 1189 TGGTGATT 1182
|||||

RESULT 6
LOCUS AA062861 377 bp mRNA EST 02-FEB-1997
DEFINITION zf70el0.s1 Soares pineal gland N3HPG Homo sapiens cDNA clone 382314
3'.
ACCESSION AA062861
NID G1557362
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 377)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le.M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
WashU-Merck EST Project
Unpublished (1995)

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 613 Std Error: 0.00
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 333.
Location/Qualifiers
1..377
/organism="Homo sapiens"
/notes="Organ: pineal gland; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
```


1st strand cDNA was primed with a Not I - oligo(dT) primer
 [5' TGTTACCAATCTGAAGTGGAGCGCCGCGCTTTTCTTTTCTTTT
 3'], double-stranded cDNA was size selected, ligated to
 Eco RI adapters (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of a modified pT73
 vector (Pharmacia). Library constructed by Bento Soares
 and M.Fatima Bonaldo."

/clone="382314"
 /clone_lib="Soares pineal gland N3HPG"
 /lab_host="DH10B (ampicillin resistant)"
 complement(<1.>377)

BASE COUNT 113 a 74 c 74 g 116 t
 ORIGIN

Query Match 22.1%; Score 345; DB 30; Length 377;
 Best Local Similarity 99.2%; Pred. No. 0.00e+00;
 Matches 358; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

Db 1 ggtctccatacatatttatgtattattattattataataataatcattcattggaagca 60
 |||||
 Cp 1545 GGTCTCTCCCATACATTTTAAATATGTATTATTAAATCAACATCATTCATAGAAAGCA 1486
 |||||

Db 61 tattacatacatgtttatatacagaattacatttttttaataaaatgtatcacagtgg 120
 |||||

Cp 1485 TATTACATACATGTTTATACATAGCAATTCATTTTAAATCAACATCATTCATAGAAAGCA 1426
 |||||

Db 121 ggcactgtttgtggaagcgttgagcttttttttaaatgagtttagagctattagataacc 180
 |||||

Cp 1425 GGCACCTGTTTGTGGAAGCGTTGGAGCTTTTAAAGAGTTTATAGACTATTAGATAACC 1366
 |||||

Db 181 actgagtaaggttaactatgtacacacaaagtgtgcatccaagagagcagcagca 240
 |||||

Cp 1365 ACTGAGTTAAAGGTAACTATGTACACACAAAGTGTGCATCCAGAGGCATATCAGCAGCA 1306
 |||||

Db 241 gaagctcttaaaagcctgtcacacagaagaagaatcacctcttccctgttggaatc 300
 |||||

Cp 1305 GAAGCTTTAAAGCCTGTACACAGGAAGAAAGATGCATCCCTTGGCTGTGGCAATC 1246
 |||||

Db 301 atttcccttagaaacaggccagcttcacctgggcccctcctgtgcccctttcaaggctg 360
 |||||

Cp 1345 ATTTCCTTTAGAAACAGGCCACCTTCACCTGGGCACCC-TGCTGCC-TTTCAGGCGTG 1188
 |||||

Db 361 g 361
 |||||

Cp 1187 g 1187

RESULT 7
 LOCUS AA062853 312 bp mRNA EST 02-FEB-1997
 DEFINITION zf70d09.r1 Soares pineal gland N3HPG Homo sapiens cDNA clone 382289
 5'

ACCESSION AA062853
 NID g1557354
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 312)
 AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
 Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
 Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
 Waterston,R., Williamson,A., Wohlmann,P. and Willson,R.
 WashU-Merck EST Project
 Unpublished (1995)

TITLE WashU-Merck EST Project
 JOURNAL
 COMMENT

Contact: Willson RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 807 Std Error: 0.00
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 213.
 Location/Qualifiers

FEATURES
 source

1..312
 /organism="Homo sapiens"
 /note="Organ: pineal gland; Vector: pT73D (Pharmacia)
 with a modified polylinker; Site.1: Not I; Site.2: Eco RI;
 1st strand cDNA was primed with a Not I - oligo(dT) primer
 [5' TGTTACCAATCTGAAGTGGAGCGCCGCGCTTTTCTTTTCTTTT
 3'], double-stranded cDNA was size selected, ligated to
 Eco RI adapters (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of a modified pT73
 vector (Pharmacia). Library constructed by Bento Soares
 and M.Fatima Bonaldo."

/clone="382289"
 /clone_lib="Soares pineal gland N3HPG"
 /lab_host="DH10B (ampicillin resistant)"

<1.>312
 BASE COUNT 93 a 67 c 56 g 94 t 2 others
 ORIGIN

Query Match 18.3%; Score 286; DB 30; Length 312;
 Best Local Similarity 97.8%; Pred. No. 0.00e+00;
 Matches 306; Conservative 0; Mismatches 4; Indels 3; Gaps 3;

Db 1 atcacagcctttaaaggcagcaggtgccccaggtgaagctggtttcttaaaagaa 60
 |||||

QY 1183 ATCACCAGCCTTGAAAGGCGCAGCGGTGCCAGGTGAAGCTGGCCTGTTTCTTAAAGGAA 1242
 |||||

Db 61 aatgattgcacaaagn-aagagatgcattcttctcctgtgtacaaagcctttaagac 119
 |||||

QY 1243 AATGATTGCCACAGGCAAGAGGATGCATCTTCTTCTCGTGTACAGGCTTTAAAGAC 1302
 |||||

Db 120 ttctgtctgtctatgcctcttggatgcacatttctgtgtacatagttaccctttaact 179
 |||||

QY 1303 TTCTGCTGCTGATATGCTCTTGGATGCACACTTTGTGTGTACATATTACCTTTAACTC 1362
 |||||

Db 180 agtggttatctaagctctaaactcattaaaaaacccaagccttccaccacaaacagt 239
 |||||

QY 1363 AGTGGTTATCTAATAGCTTAAACTCATTTAAAAAACTCCCAAGCCTTCCACCAAAACAGT 1422
 |||||

Db 240 gccccacctgtacatttttataaaaaaatgaatgcttataataaacaangtctgt 299
 |||||

QY 1423 GCCCCACCTGTATACATTTTATATAAAAAATGTAATGCTTAATGATATAAACA-TGTAATG 1481
 |||||

Db 300 aataatgctttct 312
 |||||

QY 1482 AATA-TGCTTTCT 1493

RESULT 8
 LOCUS AA107326 385 bp mRNA EST 04-FEB-1997
 DEFINITION ml95b10.r1 StrataGene mouse kidney (#937315) Mus musculus cDNA
 clone 519739 5' similar to TR:G1161100 G1161100 HYDROGEN
 PEROXIDE-INDUCIBLE PROTEIN ;

ACCESSION AA107326
 NID g1659299
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Mus.

REFERENCE 1 (bases 1 to 385)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.

TITLE The WashU-HHMI Mouse EST Project


```

LOCUS      RA210789      542 bp      mRNA      EST      14-FEB-1997
DEFINITION zfs0e08.r1 Soares NBHTGBC Homo sapiens cDNA clone 682982 5'.
ACCESSION  AF210789
NID        g1809443
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Euthera; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE  1 (bases 1 to 542)
AUTHORS    Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevisan,E., Waterston,R., Williamson,A., Wohlmann,P. and
            Wilson,R.
TITLE      The WashU-Merck EST Project
JOURNAL    Unpublished (1995)
COMMENT    Contact: Wilton RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: estewatson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28ml3 rev2 ET from Amersham
            High quality sequence stop: 414.

FEATURES   Location/Qualifiers
            source          1..542
                        /organism="Homo sapiens"
                        /note="vector: p7T73D-Pac (Pharmacia) with a modified
                        polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand
                        was prepared from germinal B-cells (flow-sorted from
                        tonsils) provided by Dr. Louis Staudt of the NCI, and was
                        then primed with a Not I - oligo(dT) primer [5'-
                        TGTACCAATCTGAAGTGGGAGCGCGCTCATTTTTTTTTTTTTTTT 3'] .
                        Double-stranded cDNA was ligated to Eco RI adaptors
                        (Pharmacia), digested with Not I and cloned into the Not I
                        and Eco RI sites of the modified pT7n3 vector. Library
                        went through one round of normalization, and was
                        constructed by Bento Soares and M. Fatima Bonaldo."
                        /clone="682982"
                        /clone_lib="Soares NBHTGBC"
                        /tissue_type="Germinal B-cell"
                        /lab_host="DH10B"
                        /lab_host="DH10B"
            mRNA            162 a 97 c 117 g 164 t 2 others
            BASE COUNT      <1...>542
            ORIGIN
            Query Match      14.7%; Score 229; DB 54; Length 542;
            Best Local Similarity 95.4%; Pred. No. 4,23e-296;
            Matches 308; Conservative 0; Mismatches 2; Indels 13; Gaps 13;

            Db 220 ttgggttcctccatcacattttaaattgtattattattttaaatacaaacatcattcatagaaa 279
            Cp 1548 TTGTGTTCTCCATACATATTTAATATGTATTATATTTTAAATCAACATCATTCATAGAAA 1489
            Db 280 gcattattacatcatgtttatcacataaacattacatttttttaataaaaaatgtatacagg 339
            Cp 1488 GCATATTACATACATGTTTATACATAAAGCATTTATTTTAAATAAAATGTATACAGG 1429
            Db 340 tggggcactgttttgggtgaaggcttgagtttttttaataagtttagagctattaagat 399
            Cp 1428 TGGGGCACTGTTTGGTGGAGGCTTGAGATTTTTTTT-AAAGAGTTTAAAGCTATTAGAT 1370
            Db 400 aaccactgagttaaagggttaactatggtacacacaaaagggtgtgccatccaagaaggcata 459
            Cp 1369 AACCACTGAGTTAAAGG-TAACTATG-TACACACAAAG-TGTGC-ATCCAAGA-GGCATA 1315
            Db 460 gcagcagcagaagtcctttaaggcttgaatcncacagggaagaaggatgcatoctcctt 519
            |||||||

```

```

Cp 1314 TCAGCAGCAGAA-GTCTTTAAAGCTTTGTAT-ACCAGG-AAGAAAG-ATGCATCCTC-TT 1260
Db 520 ggccttggtggaatcattttcc 542
Cp 1259 G-CCTTG-TGGCAATCATTTTCC 1239

RESULT 11
LOCUS    AA152779      411 bp      mRNA      EST      11-FEB-1997
DEFINITION mr78c03.r1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
            603556 5' similar to TR.G1161100 G1161100 HYDROGEN
            PEROXIDE-INDUCIBLE PROTEIN ;.
ACCESSION AA152779
NID        g1724432
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE  1 (bases 1 to 411)
AUTHORS    Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE      The WashU-HMI Mouse EST Project
JOURNAL    Unpublished (1996)
COMMENT    Contact: Marra M/Mouse EST Project
            WashU-HMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:368988
            Possible reversed clone: similarity on wrong strand
            Seq primer: -28ml3 rev1 ET from Amersham
            High quality sequence stop: 207.

FEATURES   Location/Qualifiers
            source          1..411
                        /organism="Mus musculus"
                        /strain="NIH/Swiss"
                        /note="Vector: phuescript SK-; Site.1: EcoRI; Site.2:
                        XhoI; Cloned unidirectionally. Primer: Oligo dT. 93
                        pooled NIH/Swiss 13 day embryo hearts. Average insert
                        size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
                        GAATTCGGCAGCAG 3' -3' adaptor sequence: 5'
                        CTCGAGTTTCTTTTCTTTTCTTTT 3'"
                        /clone="603556"
                        /clone_lib="Stratagene mouse heart (#937316)"
                        /sex="pooled"
                        /dev_stage="13 day embryos"
                        /lab_host="SOLR (kanamycin resistant)"
                        /lab_host="SOLR (kanamycin resistant)"
            mRNA            75 a 122 c 131 g 83 t
            BASE COUNT      <1...>411
            ORIGIN
            Query Match      7.9%; Score 123; DB 50; Length 411;
            Best Local Similarity 91.8%; Pred. No. 5.49e-134;
            Matches 135; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

            Db 264 ccagccggtgacctgttcgacaaagacgcggtccgtacgtgcccctcgatgtattg 323
            Cp 329 CCGGGCAGAGACCATGTTTGACAAAGACGCGGCTCGCTACGTGCGCCCTCGATGTGCTG 388
            Db 324 cgttttctggtgagtgaccttttggaaattcttacttcaaggcataccccctttcagcg 383
            |||||||
            Cp 389 CGTGTGCTGGCTGGATTGCTCTTTTGCAATTTCTTACTTCAAGCATACCCCTCCACAG 448
            |||||||

```

```

Db 384 aggaatattctgtaagtgaactccat 410
||||| ||||||| ||||||| |||||||
QY 449 AGGAGTATTCTGTAATGATGAGTCCAT 475

RESULT 12
ID HS368337 standard; RNA; EST; 241 BP.
AC W20368;
NI 91296266
DT 05-MAY-1996 (Rel. 47, Created)
DT 05-MAR-1997 (Rel. 51, Last updated, Version 2)
DE zc68b11.s1 Soares fetal heart NBHHL19W Homo sapiens cDNA clone
DE 327453 3'.
KW EST.
OS Homo sapiens (human)
OC Eukaryotae; mitochondria eukaryotes; Metazoa; Chordata;
OC Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP 1-241
RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,
RA Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,
RA Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F.,
RA Treveskis E., Waterston R., Williamson A., Wohlmann P., Wilson R.;
RT "The WashU-Merck EST Project";
RL Unpublished.
CC Contact: Wilson RK WashU-Merck EST Project Washington University
CC School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,
CC MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:
CC estewatson.wustl.edu This clone is available royalty-free through
CC LLNL; contact the IMAGE Consortium (infoimage.llnl.gov) for
CC further information. Insert Length: 981 Std Error: 0.00 Seq primer:
CC mob.REGA+ET High quality sequence stop: 191.
FH Key
FH Location/Qualifiers
FT 1..241
FT source
FT /organism="Homo sapiens"
FT /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
FT modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
FT strand cDNA was primed with a Not I - oligo(dT) primer [5'
FT TGTACCAATCGAAGTGGAGCGCGCATCTTTTCTTTTCTTTTCTTTT 3']
FT double-stranded cDNAs size selected, ligated to Eco RI
FT adapters (Pharmacia), digested with Not I and cloned into
FT the Not I and Eco RI sites of a modified pT7T3 vector
FT (Pharmacia). Library went through one round of
FT normalization to a Cot = 5. Library constructed by M.Fatim
a
FT Bernaldo. This library was constructed from the same fetus
FT as the fetal lung library, Soares fetal lung NBHHL19W."
FT /clone="327453"
FT /db="GenBank"
FT /sex="unknown"
FT /dev_stage="19 weeks"
FT /lab_host="PH10B (ampicillin resistant)"
FT complement(1..241)
FT mRNA
SQ Sequence 241 BP; 87 A; 40 C; 36 G; 76 T; 2 other;

Query Match 7.7%; Score 120; DB 87; Length 241;
Best Local Similarity 99.2%; Pred.No. 1.56e-129;
Matches 126; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 1 gccaccctgtatcacattttttattataaaatngtaatgcttatgtataaactgtatgt 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1423 GCCCCACCTGTATACATTTTATTAAAAAAT-GTAATGCTTATGTATAAACATGTATGT 1481

Db 61 aatagctttctatgaatgattgtttgatttaataataataacattataaactgtatggag 120
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1482 AATATGCTTCTATGAATGATGTTGTTGTTAAATATAATATAATATAATATAATGATGGAG 1541
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 121 aaccaa 127
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1542 AACCAA 1548
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

RESULT 13
 LOCUS human STS WI-15758.
 ACCESSION G23282
 NID 91343608
 KEYWORDS STS sequence; primer; sequence tagged site.
 SOURCE human STSs derived from sequences in dbEST and the Unigene collection.
 ORGANISM Homo sapiens
 Eukaryotae; mitochondria eukaryotes; Metazoa; Chordata;
 Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 501)
 AUTHORS Hudson,T.
 TITLE Whitehead Institute/MIT Center for Genome Research; Physically Mapped STSs
 JOURNAL Unpublished (1995)
 COMMENT
 Contact: Thomas Hudson
 Whitehead Institute/MIT Center for Genome Research
 Whitehead Institute for Biomedical Research
 9 Cambridge Center, Cambridge MA 02142 USA
 Tel: 617 352 1900
 Fax: 617 352 1902
 Email: thudson@genome.wi.mit.edu
 Primer A: TATTACGATTGGCAAAATGAAGC
 Primer B: TACTTTCAGGTGTACAGATTAAACA
 STS size: 129
 PCR Profile:
 Presoak:
 Denaturation:
 Annealing: 56 degrees C
 Polymerization:
 PCR Cycles: 35
 Thermal Cycler:
 Protocol:
 Template: 10 ng
 Primer: each 5 pM
 dNTPs: each 4 nM
 Taq Polymerase: 0.025 units/ul
 Total Vol: 20 ul
 Buffer:
 MgCl2: 1.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 9.3
 Derived from dbEST (genbank accession R45479).
 Location/Qualifiers
 source
 1..501
 /organism="Homo sapiens"
 /map="188.1 cR from top of Chr5 linkage group"
 STS
 primer_bind 23..151
 primer_bind 23..46
 BASE COUNT 133 a 88 c 117 g 157 t 6 others
 ORIGIN
 Query Match 7.6%; Score 119; DB 43; Length 501;
 Best Local Similarity 84.9%; Pred.No. 4.73e-128;
 Matches 202; Conservative 0; Mismatches 26; Indels 10; Gaps 10;
 Db 255 gctgtatgtatgcgggtttcttcaaggaggaactctttttaaaggagaaagagagg 314
 ||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 QY 1068 GCTGTATGTATCGGATTCTCTCAAGAAAGAACTCTCTTTTAAAG-AAAG-AAAGAGG 1125
 ||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 315 ggggggtctcacaactctgcctggaacacccacacacacacacacacacacacac 374
 ||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 QY 1126 AGGA-CTCTCATCACTCTG-CATG-AAACACCAACTGGGAAT-CACATATCGAG- 1180
 ||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 375 caatcacccagccttgaaaggcggggtgcccgggggtgagggcggtggtttttcttaa 434
 ||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY	1181	CAATCACCAGCCTTAAGG-CACGAGGGTCCCAGG-TGAAGCTGG-CCTGTTTCTTAA	1237
Db	435	ggggaataatggtgcacaggnacagggggtgcctcttcttccctccggggtacaggttt	492
QY	1238	AGGAAATGATTGCCACAGGACAGAGGATGCATCTTCTCTCTGTTGTACAGCCCTT	1295
RESULT	14		
ID	MM1155742	standard; RNA; EST; 313 BP.	
AC	AA237956;		
NI	g1862038		
DT	06-MAR-1997 (Rel. 51, Created)		
DT	06-MAR-1997 (Rel. 51, Last updated, Version 1)		
DE	mx78901.r1 Soares mouse NML Mus musculus cDNA clone 692496 5'		
DE	similar to TR:G1161100 G1161100 HYDROGEN PEROXIDE-INDUCIBLE PROTEIN		
DE	;		
KE	Mus musculus (house mouse)		
OS	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;		
OC	Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
OC	Mus.		
RN	[1]		
RP	1-313		
RA	Marra M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T., Geisel S., Kucaba T., Lacy M., Le M., Martin J., Morris M., Schellenberg K., Steptoe M., Tan F., Underwood K., Moore B., Theising B., Wylie T., Lennon G., Soares B., Wilson R., Waterston R.;		
RT	"The WashU-HMI Mouse EST Project";		
RL	Unpublished.		
CC	Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project		
CC	Washington University School of Medicine 4444 Forest Park Parkway,		
CC	Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810		
CC	Email: mouseest@wustl.edu This clone is available		
CC	royalty-free through LLNL; contact the IMAGE Consortium		
CC	(info@image.llnl.gov) for further information. MGI:426056 Possible		
CC	reversed clone: similarity on wrong strand Seq primer: -28m13 rev2		
CC	ET from Amersham High quality sequence stop: 57.		
CC	Key Location/Qualifiers		
FT	Source		
FT	1..313		
FT	/organism="Mus musculus"		
FT	/note="Vector: pT7T3D-Pac (Pharmacia) with a modified		
FT	polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA		
FT	was primed with a Not I - oligo(gt) primer [5,		
FT	TGTTACCAATCTCAAGTGGAGCGCGCGCAACTCTTTTCTTTT 3'];		
FT	double-stranded cDNA was ligated to Eco RI adaptors		
FT	(Pharmacia), digested with Not I and cloned into the Not I		
FT	and Eco RI sites of the modified pT7T3 vector. Library		
FT	constructed and normalized by Bento Soares and M.Patima		
FT	Bonaldo."		
FT	/clone="692496"		
FT	/clone_lib="Soares mouse NML"		
FT	/tissue_type="Liver"		
FT	/lab_host="DH10B"		
FT	<1..>313		
FT	mRNA		
SQ	Sequence 313 BP; 83 A; 58 G; 87 T; 0 other;		
	Query Match 5.6%; Score 88; DB 91; Length 313;		
	Best Local Similarity 66.5%; Pred. No. 6.59e-83;		
	Matches 208; Conservative 0; Mismatches 102; Indels 3; Gaps 3;		
Db	3	gagtcacactgtacgtccactccagtggtgtttagtcacacactcagaaagcc-ac	61
Cp	1037	GAGTCCAGTCAACATCGCTCCAGTGGTGTTTATATACAAACTCGAGAAAGCCCCAC	978
Db	62	atataatgaaaaagcaatgagccaaactgagcaatggtcgttaagaagcctggccaatt	121
Cp	977	ATAAATGATACGCGCAACACAGCAACAAATTCGAGTGTGGCGCTRAGA-GTCTGCCCACT	919
Db	122	tcccttaactcgtgcttgaggttaagtctgcgaacacaggtttagttacttgggattg	181
Cp	918	CTCCCTTCATCCTGCTTGAAGATAAAGTGCACAAACAGCAGTACAGTACATGGAACG	859
Db	182	agaatttccgggttggaagccaaactggccttcttcttgattccctt-caatcccttggc	240
Cp	858	AAGAGTGGCCTGATAGAGGACAACTGCTTCTTAACTCTTTCGCAATCCCTCGAC	799
Db	241	ataaggagttcttcattatggccattaatgagttggatttgaacagctctcgggtataga	300
Cp	798	ATATGTAGTATTCAATGTAACCATCGCTCGAGTGTGATTTTGACCAATCTGGATCACA	739
Db	301	gatgccagaaatg 313	
Cp	738	CATCCAAGAAGTG 726	
RESULT	15		
LOCUS	AA237956	313 bp	MRNA
DEFINITION	mx78901.r1 Soares mouse NML Mus musculus cDNA clone 692496 5'		
	similar to TR:G1161100 G1161100 HYDROGEN PEROXIDE-INDUCIBLE PROTEIN		
	;		
ACCESSION	AA237956		
NID	g1862038		
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;		
	Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
	Mus.		
REFERENCE	1 (bases 1 to 313)		
AUTHORS	Marra M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T., Geisel S., Kucaba T., Lacy M., Le M., Martin J., Morris M., Schellenberg K., Steptoe M., Tan F., Underwood K., Moore B., Theising B., Wylie T., Lennon G., Soares B., Wilson R. and Waterston R.		
TITLE	The WashU-HMI Mouse EST Project		
JOURNAL	Unpublished (1996)		
COMMENT	Contact: Marra M/Mouse EST Project		
	WashU-HMI Mouse EST Project		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: mouseest@wustl.edu		
	This clone is available royalty-free through LLNL; contact the		
	IMAGE Consortium (info@image.llnl.gov) for further information.		
	MGI:426056		
	Possible reversed clone: similarity on wrong strand		
	Seq primer: -28m13 rev2 ET from Amersham		
	High quality sequence stop: 57.		
FEATURES	Location/Qualifiers		
Source	1..313		
	/organism="Mus musculus"		
	/note="Vector: pT7T3D-Pac (Pharmacia) with a modified		
	polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA		
	was primed with a Not I - oligo(gt) primer [5,		
	TGTTACCAATCTCAAGTGGAGCGCGCGCAACTCTTTTCTTTT 3'];		
	double-stranded cDNA was ligated to Eco RI adaptors		
	(Pharmacia), digested with Not I and cloned into the Not I		
	and Eco RI sites of the modified pT7T3 vector. Library		
	constructed and normalized by Bento Soares and M.Patima		
	Bonaldo."		
	/clone="692496"		
	/clone_lib="Soares mouse NML"		
	/tissue_type="Liver"		
	/lab_host="DH10B"		
	<1..>313		
	mRNA		
	BASE COUNT 83 a 58 c 85 g 87 t		
	ORIGIN		
	Query Match 5.6%; Score 88; DB 66; Length 313;		
	Best Local Similarity 66.5%; Pred. No. 6.59e-83;		
	Matches 208; Conservative 0; Mismatches 102; Indels 3; Gaps 3;		


```

Db      239 tvhylllyktgvanwltelelfaI 261
        :: :|:: :: : ||:: ::|
QY      101 AIGTFLEGAASQSOLDIAKYSI 123

```

RESULT 2
ID W11257 standard; Protein; 564 AA.

DT 17-MAR-1997 (first entry)
DE Heart calcium/calmodulin dependent phosphodiesterase.
KW Cyclic GMP stimulated phosphodiesterase; CGS-PDE; bovine; brain; cAMP;
KW adrenal gland; transmembrane signal; extracellular hormone; antibody;
KW neurotransmitter; cGMP; calcium/calmodulin dependent phosphodiesterase;
KW CaM-PDE; human.

PN	US5380771-A.	
PP	03-DEC-1996.	
PD	19-APR-1991;	688356.
PF	19-APR-1991;	PR
PR	19-APR-1991;	US-688356.
PR	20-APR-1992;	US-872844.
PR	29-AUG-1994;	US-297494.
PR	(UNTW) UNIV WASHINGTON.	
PI	Beavo JA, Charbonneau H,	
PI	WPI: 97-033573/03.	
DR	N-PDSR: T51117	Sonnen

DNA encoding cGMP-stimulated phosphodiesterase - for prodn. of
 recombinant enzyme
 Example 7: Column 117-122; 68pp; English.
 This sequence represents the calcium/calmodulin dependent
 phosphodiesterase (CaM-PDE) encoded by the cDNA insert of pcamHella.
 This cDNA insert was isolated from human heart mRNA. The cyclic
 nucleotide phosphodiesterases (PDEs) catalyze the hydrolysis of 3', 5'
 cyclic nucleotides, such as cAMP and cGMP, to their corresponding
 5'-nucleotide monophosphates. The PDEs are therefore important in the
 control of the cellular concentration of cyclic nucleotides. The PDEs
 are, in turn, regulated by transmembrane signals or second messenger
 ligands such as calcium ion or cGMP. The PDEs therefore have a central
 role in regulating the flow of information from extracellular hormones,
 neurotransmitters, or other signals that use the cyclic nucleotides as
 messengers. PDEs are present in most of the cells and tissues of
 eukaryotic organisms, but only in trace amounts. CaM-PDEs are responsive
 to intracellular calcium, which leads to a decreased intracellular
 concentration of cAMP, and/or cGMP. The cDNA encoding this sequence can
 be used for the production of recombinant CaM-PDE, which may have
 therapeutic and diagnostic uses. This protein may also be useful for
 diagnostic antibody production.
 Sequence 564 AA;

Query Match	4.88;	Score 102;	DB 20;	Length 564;
Best Local Similarity	24.18;	Pred. No. 6.07e+00;		
Matches	20;	Conservative	27;	Mismatches 32;
				Indels 4;
				Gaps 3;

```

Db      37 dhalkfifyelltrydlisrfrkkipalsalvsfv-ealevgysk--hknpynhlmhaadtq 93
      |:::| |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY     42 DESIKYPKYKEDTYPYALGGIIIFSIIVILGETLSV-YCNLLHSFSIRNNYIATYK 100

```

```
Db      94 tvhylllyktgtvanwltelelfaI 116  
       :: :|:: :: : ||:: :: |  
QY     101 AIGTFLFGAAASQSLTDIAKYSI 123
```

RESULT 3
ID R69733 standard; Protein: 564 AA.

DT	11-OCT-1995 (first entry)
DE	Cyclic-GMP stimulated nucleotide PDE clone pcam-Hella.
KW	Cyclic-GMP stimulated nucleotide phosphodiesterase; antibodies;
KW	human; hormones; neurotransmitters; transmission regulation;
KW	enzyme purification; clone pcam-Hella.

OS	Homo sapiens.
PN	US5389527-A.
PD	14-FEB-1995.
PF	19-APR-1991; 688356.

PD 14-FEB-1995.
PF 19-APR-1991; 688356.

PR 19-APR-1991; US-688356.
PR 20-APR-1992; US-872844.
PA (UNIW) UNIV WASHINGTON.
PI Beavo JA, Charbonneau H, Sonnenburg WK;
DR WPI: 95-090205/12.
DR N-PSDB; Q83982.
PT New nucleic acid encoding cyclic-GMP stimulated nucleotide
PT phospho-di-esterase - and related vectors and transformed cells,
PT useful for screening cpds. for phospho-di-esterase modulating
PT activity
PS Example 7; Columns 117-122; 69pp; English.
CC Q83982 encodes R9733 the human cyclic-GMP stimulated nucleotide
CC phospho-di-esterase (Cam PDE) clone pcam-Hella. Eukaryotic cells
CC that express Cam PDE can be used to screen cpds. for the ability to
CC modulate Cam PDE activity. Cam PDEs are involved in regulating
CC the transmission of information from hormones, neurotransmitters
CC or other systems that use cyclic nucleotides as messengers.
CC Antibodies raised against Cam PDE can be used for enzyme purificn.,
CC or determination.
SQ Sequence 564 AA.

Query Match 4.8%; Score 102; DB 13; Length 564;
Best Local Similarity 24.1%; Pred. No. 6.07e+00;
Matches 20; Conservative 27; Mismatches 32; Indels 4; Gaps 3;

```

Db 37 dbalkfifvellttrydlisrfkipisalvsfv-ealevgysk--hknpyhnlmhoadvtq 93
   |:::| | |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 42 DESIKYPKEDTIPYALGGIIPESIIIVILGETLSV-YCNLHLSNSIRNNYATIK 100

```

Db 94 tvhylllyktgvnwltelelfaI 116
 :: :| :: : : || :: : : |
QY 101 AIGTFLEGAASQSLTDIKYSI 123

RESULT 4
ID R28413 standard: Protein: 564 AA.

AC	R28413:
DE	19-MAR-1993 (first entry)
DE	Prod. of cDNA insert of pcamHella.
KW	Calcin/ calmodulin; stimulated.
KW	cyclic; nucleotide; bovine:
KW	phosphodiesterase.
OS	Homo sapiens.
PN	W09218541-A.
PD	29-OCT-1992.
PF	20-APR-1992; U03222.
PR	19-APR-1991; US-688356.
PA	(UNIW) UNIV WASHINGTON.
PI	Beavo JA, Bentley JK, Charbonneau H, Sonnenburg WK;
DR	WPI: 92-382051/46.
DR	N-PDSB: Q30188.
DR	

New DNA encoding mammalian cyclic nucleotide phospho-di-esterase - and derived vectors and host cells, useful for screening cpds. for inhibitory or activating activity

Example 7: Page 114; 133pp; English.

An approx. 2.4 kb fragment of pcamH3EF was used to screen a human heart cDNA library to yield two plaques which hybridised to the pcamH3EF probe. The Bluescript SK- plasmid pcamHella was excised *in vivo* from one of these clones. Sequence analysis of pcamHella showed that the insert began at nucleotide 610 of pcamH3EF and was nearly identical to nucleotide 2066, at which point the DNA sequence diverged from that of pcamH3EF. The cDNA insert of pcamHella continued for ca. 0.6 kb. The consequence of this divergence is to alter the carboxy terminus of the protein encoded by the ORF. pcamH3EF cDNA can encode a protein of 634 residues, and assuming the 5' end of pcamHella cDNA is the same as that of the 5' end of pcamH3EF (5' to nucleotide 610), pcamHella can encode a 709 residue protein. These divergent 3' ends may be the consequence of alternative splicing, lack of splicing, or unrelated DNA sequences being juxtaposed during the cloning process.

Sequence 564 AA:

Query Match 4.8%; Score 102; DB 5; Length 564;
Best Local Similarity 24.1%; Pred. No. 6.07e+00;

RESULT	6
ID	R69732 standard; Protein; 634 AA.
AC	R69732;
DT	11-OCT-1995 (first entry)
DE	Cyclic-GMP stimulated nucleotide PDE clone pcamH3EF.
KW	Cyclic-GMP stimulated nucleotide phospho-diesterase; antibodies;

```

RESULT      7
ID      R73055 standard; Protein; 1873 AA.
AC      R73055;
DE      02-NOV-1995 (first entry)
DT      Rabbt skeletal calcium channel (alpha)1-subunit.
KW      Calcium channel; (alpha)1 subunit.
OS      Oryctolagus cuniculus.
PH      key      Location/Qualifiers
FT      Region      950..1100
FT      /label= IIF7 monoclonal antibody epitope
FT      /label= Misc_difference 1808
FT      /label= Thr
FT      /note= "In Tanabe et al"
FT      /label= Ala
FT      /label= Misc_difference 1815
FT      /note= "In Tanabe et al"
FT      /label= Ala
FT      /label= Misc_difference 1835
FT      /note= "In Tanabe et al"
FT      /label= Ala
FT      /label= Modified_site 79
FT      /label= Potential N-glycosylation site
FT      /label= Modified_site 257
FT      /label= see above
FT      /label= Modified_site 797
FT      /label= see above
FT      /label= Modified_site 1464
FT      /label= see above
FT      /label= Modified_site 1674
FT      /label= see above
FT      /label= Modified_site 687
FT      /label= potential cAMP-dependent phosphorylation
FT      /label= Modified_site 1502
FT      /label= see above

```

FT Modified_site 1575
 FT /label= see above
 FT Modified_site 1757
 FT /label= see above
 FT Modified_site 1772
 FT /label= see above
 FT Modified_site 1854
 FT /label= see above
 FT Modified_site 1552
 FT /label= see above
 FT Region 52..70
 FT /label= transmembrane region
 FT Protein 89..108
 FT /label= see above
 FT Region 121..139
 FT /label= see above
 FT Region 161..179
 FT /label= see above
 FT Region 199..218
 FT /label= see above
 FT Region 310..334
 FT /label= see above
 FT Region 433..451
 FT /label= see above
 FT Region 467..486
 FT /label= see above
 FT Region 495..513
 FT /label= see above
 FT Region 524..542
 FT /label= see above
 FT Region 562..581
 FT /label= see above
 FT Region 637..661
 FT /label= see above
 FT Region 800..818
 FT /label= see above
 FT Region 835..854
 FT /label= see above
 FT Region 867..885
 FT /label= see above
 FT Region 893..912
 FT /label= see above
 FT Region 931..950
 FT /label= see above
 FT Region 1041..1065
 FT /label= see above
 FT Region 1119..1137
 FT /label= see above
 FT Region 1153..1172
 FT /label= see above
 FT Region 1181..1199
 FT /label= see above
 FT Region 1232..1250
 FT /label= see above
 FT Region 1270..1289
 FT /label= see above
 FT US5407820-A.
 PD 18-APR-1995.
 PF 04-APR-1988; 176899.
 PR 04-APR-1988; US-176899.
 PR 04-APR-1989; US-603751.
 PR 13-JUL-1992; US-914231.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Brenner R, Ellis SB, Harpold MM, Schwartz A, Williams ME;
 DR WPI: 95-161088/21.
 DR N-PSDB; Q87978.
 PT DNA encoding alpha-1 and alpha-2 calcium channel sub-unit(s) -
 PT for detecting calcium channel agonists/antagonists or
 PT Lambert-Eaton syndrome.
 PS Disclosure: Fig 1: 32pp; English.
 CC The cDNA sequence Q87978 is consistent with an approx 6,500 nt DHP
 receptor (alpha1) mRNA. It is 94% identical to the cDNA sequence
 CC encoding the DHP receptor reported by Tanabe et al., Nature, 328:313-

CC 318, 1987. Nt differences were identified at 33 posns, of which
 CC three (see FT) also result in AA changes (see R73055 FT). The
 CC deduced AA sequence yields a calculated Mr of 212,143, in contrast to
 CC the observed Mr 155K-170K, previously reported by others using SDS
 CC polyacrylamide gel electrophoresis. The AA sequence is 99.8%
 CC identical to that described by Tanabe et al. It contains four
 CC internal repeated sequence regions. It has been proposed that the
 CC segments of the four internally repeated regions represent 24
 CC transmembrane segments. A eukaryotic cell expressing a chimeric
 CC cDNA is claimed, in which the first cDNA is selected from the gp.
 CC consisting of cDNA that encodes a protein that has the AA sequence
 CC in R73055 but with the Tanabe et al substitutions indicated in FT.
 SQ Sequence 1873 AA;
 Query Match 4.8%; Score 102; DB 13; Length 1873;
 Best Local Similarity 23.8%; Pred. No. 6.07e+00;
 Matches 20; Conservative 22; Mismatches 39; Indels 3; Gaps 3;
 Db 821 piraesvnrnqilgyfdiaftsvftveivkmtygaflhkgscfcrnyfnildllvvavsl 880
 QY 48 PYKEDTIPYALLGGIIPF-SIIIVILGETLSVYCNLLHNSFIRNNYIATYIKAIQTF 106
 Db 881 ismglesstsvkvkrlrvlrp 904
 QY 107 FGAASQSLTDIAKY-SIGR-LRP 128
 RESULT 8
 ID P95645 standard; protein; 1873 AA.
 AC P95645;
 DT 21-MAR-1990 (first entry)
 DE Rabbit skeletal muscle alpha-1 sub-unit gene product.
 KW Skeletal muscle
 OS Sylvilagus (rabbit).
 PN W08909834-A.
 PD 19-OCT-1989.
 PF 04-APR-1988; 001408.
 PR 04-APR-1988; US-176899.
 PA (SALK) Salk Inst for Biol Stud.
 PI Ellis SB, Williams ME, Harpold MM, Schwartz A, Sartor J;
 DR WPI: 89-324236/44.
 DR N-PSDB; N91778.
 PT New DNA encoding alpha-2 subunit of animal calcium channel - also new
 PT protein product and eukaryotic cells for testing cpds. for calcium
 PT agonist or antagonist activity
 PS Disclosure: page 16-1 to 18-3; 68pp; English.
 CC Also used to diagnose Lambert-Eaton syndrome by reacting test serum
 CC with alpha-1 and alpha-2 subunits. Labelled fragments can be used as
 CC probes.
 SQ Sequence 1873 AA;
 Query Match 4.8%; Score 102; DB 13; Length 1873;
 Best Local Similarity 23.8%; Pred. No. 6.07e+00;
 Matches 20; Conservative 22; Mismatches 39; Indels 3; Gaps 3;
 Db 821 piraesvnrnqilgyfdiaftsvftveivkmtygaflhkgscfcrnyfnildllvvavsl 880
 QY 48 PYKEDTIPYALLGGIIPF-SIIIVILGETLSVYCNLLHNSFIRNNYIATYIKAIQTF 106
 Db 881 ismglesstsvkvkrlrvlrp 904
 QY 107 FGAASQSLTDIAKY-SIGR-LRP 128
 RESULT 9
 ID W04217 standard; protein; 397 AA.
 AC W04217;
 DT 23-NOV-1996 (first entry)
 DE Rat dorsal root ganglion P2x receptor.
 KW ATP P2x receptor; purinoceptor; ligand-gated ion channel; agonist;
 KW antagonist; epilepsy; cognition; emesis; pain; asthma;
 KW peripheral vascular disease; hypertension; irritable bowel syndrome;
 KW premature ejaculation; cystitis; therapy.

OS Rattus sp.
PN W09533048-A2.

PD 07-DEC-1995.
PF 24-MAY-1995; E01968.
PR 27-MAY-1994; GB-010664.
PS 09-FEB-1995; GB-002480.
PA (GLAXO) GLAXO GROUP LTD.
PI Buell GN, Valera S;
DR WPI; 96-030561/03.
DR N-PSDB; T33854.
PT DNA encoding ATP P2x receptors of the purinoceptor family - for
PT screening cpds. useful in treating epilepsy, cognition, emesis,
PT pain, asthma, peripheral vascular disease, hypertension, etc.
PS Claim 1; Fig 3; 82pp; English.

CC The amino acid sequence (W04217) of rat dorsal root ganglion P2x
CC receptor was deduced from a cDNA clone (T33854) isolated from a
CC cDNA library. P2x receptors (see also W04215-16 and W04218) are
CC ligand-gated ion channels that open upon binding of extracellular
CC ATP. Recombinant P2x receptors can be produced in prokaryotic or
CC eukaryotic, pref. mammalian, host cells. Their widespread distr.
CC throughout the body and possibly numerous physiological roles make
CC P2x receptors useful for the screening of new cpds. (agonists or
CC antagonists) for the treatment of a number of pathological states.
SQ Sequence 397 AA;

Query Match 4.5%; Score 96; DB 18; Length 397;
Best Local Similarity 31.1%; Pred. No. 1.61e+01;
Matches 14; Conservative 14; Mismatches 16; Indels 1; Gaps 1;

Db 336 tvlcdillnflkgadhykarkfevttikgtastnvpvfdsq 380

QY 239 AILVAVYVDFKERTSPKRR-EEDSHTTLHETPTTGNHPSNH 282

RESULT 10

ID W20522 standard; Protein; 186 AA.

AC W20522; 1997 (first entry)
DE H. pylori cell envelope protein 4826401.aa.
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
OS Helicobacter pylori.

PN W09640893-A1.

PD 19-DEC-1996.

PF 06-JUN-1996; U09122.

PR 07-JUN-1995; US-487032.

PR 01-APR-1996; US-630405.

PA (ASTR) ASTRA AB.

PI Berglindh OT, Smith D, Mellgaard BL;

DR WPI; 97-052306/05.

DR N-PSDB; T67677.

PT Helicobacter pylori nucleic acid sequences and related
PT polypeptides(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter

PS Claim 56; Page 682; 1481pp; English.

CC This sequence is a H. pylori cell envelope protein.

CC The protein may be used in a vaccine to prevent or treat H. pylori

CC infection or to identify H. pylori polypeptide binding compounds,

CC useful as potential H. pylori life cycle activators or inhibitors.

CC The genomic sequence of H. pylori (ATCC 55679) was determined from

CC overlapping contigs generated by mechanically shearing the bacterial

CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,

CC and the predicted coding regions defined by computer evaluation. To

CC identify likely H. pylori antigens for vaccine development, the amino

CC acid sequences predicted from various ORF were analysed for significant

CC homology to other known or exported membrane proteins. Having identified

CC and determined the sequences of interest, particular regions can be

CC isolated from H. pylori by PCR amplification for recombinant polypeptide

CC production, e.g. in E. coli hosts.

SQ Sequence 186 AA;

Query Match 4.4%; Score 93; DB 21; Length 186;

Best Local Similarity 29.1%; Pred. No. 2.59e+01;
Matches 16; Conservative 20; Mismatches 15; Indels 4; Gaps 4;

Db 31 qsiyvallgtlvtltpv-s-sl-vlkatphsefl-trflnriyaplleff 81

QY 52 DTIPYALLGGIIIPFSIIIVILGETLSVYCNLLHSNFIRNNYIATYKAIGTFL 106

RESULT 11

ID W20587 standard; protein; 254 AA.

AC W20587; 1997 (first entry)

DT 14-JUL-1997

DE H. pylori cell envelope protein, 01ae22001orf2.

KW Vaccine; prevention; treatment; infection; identification;

KW binding compound; bacterium; life cycle; activator; bacteria;

KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;

KW cell envelope.

OS Helicobacter pylori.

PN W09640893-A1.

PD 19-DEC-1996.

PF 06-JUN-1996; U09122.

PR 07-JUN-1995; US-487032.

PR 01-APR-1996; US-630405.

PA (ASTR) ASTRA AB.

PI Berglindh OT, Smith D, Mellgaard BL;

DR WPI; 97-052306/05.

DR N-PSDB; T67840.

PT Helicobacter pylori nucleic acid sequences and related

PT polypeptides(s) - useful for vaccines to treat or prevent H. pylori

PT infection, and to detect Helicobacter

PS Claim 56; Pages 1018-1019; 1481pp; English.

CC The present sequence is a Helicobacter pylori cell envelope

CC protein.

CC The protein may be used in a vaccine to prevent or treat

CC H. pylori infection or to identify H. pylori polypeptide binding

CC compounds, useful as potential H. pylori life cycle activators or

CC inhibitors. The genomic sequence of H. pylori (ATCC 55679) was

CC determined from overlapping contigs generated by mechanically

CC shearing the bacterial DNA. The sequences were analysed for ORF of

CC at least 180 nucleotides, and the predicted coding regions defined

CC by computer evaluation. To identify likely H. pylori antigens for

CC vaccine development, the amino acid sequences predicted from

CC various ORF were analysed for significant homology to other known

CC or exported membrane proteins. Having identified and determined

CC the sequences of interest, particular regions can be isolated from

CC H. pylori by PCR amplification for recombinant polypeptide

CC production, e.g. in E. coli hosts.

SQ Sequence 254 AA;

Query Match 4.4%; Score 93; DB 21; Length 254;

Best Local Similarity 29.1%; Pred. No. 2.59e+01;

Matches 16; Conservative 20; Mismatches 15; Indels 4; Gaps 4;

Db 80 qsiyvallgtlvtltpv-s-sl-vlkatphsefl-trflnriyaplleff 130

QY 52 DTIPYALLGGIIIPFSIIIVILGETLSVYCNLLHSNFIRNNYIATYKAIGTFL 106

RESULT 12

ID R52832 standard; Protein; 466 AA.

AC R52832;

DT 10-NOV-1994 (first entry)

DE Sequence of human alpha 1 C adrenergic receptor.

KW Alpha 1C; adrenergic receptor; antagonist; neuroreceptor.

OS Homo sapiens.

PN W09410989-A.

PD 26-MAY-1994.

PF 12-NOV-1993; U10950.

PR 13-NOV-1992; US-975867.

PA (SYNA-) SYNAPTIC PHARM CORP.

PI Branchek TA, Chiu G, Forray CC, Gluchowski C, Hartig PR;

PI Wetzel JW;

DR WPI; 94-183130/22.

KW extracellular matrix; cytoskeleton; heterodimer; laminin receptor;
OS Immunoprecipitation; JAR; choriocarcinoma.

PH Key Location/Qualifiers
FT Domain 1012..1037
FT /label= Transmembrane_domain
FT Peptide 1..23
FT /label= Signal_peptide
FT Protein 24..1073
FT /label= Mature_protein
FT Modified_site 223
FT /label= Glycosylation_site
FT Modified_site 284
FT /label= Glycosylation_site
FT Modified_site 370
FT /label= Glycosylation_site
FT Modified_site 513
FT /label= Glycosylation_site
FT Modified_site 731
FT /label= Glycosylation_site
FT Modified_site 748
FT /label= Glycosylation_site
FT Modified_site 891
FT /label= Glycosylation_site
FT Modified_site 927
FT /label= Glycosylation_site
FT Modified_site 958
FT /label= Glycosylation_site
FT Binding_site 230..238
FT /note= "Putative cation binding domain"
FT Binding_site 324..332
FT /note= "Putative cation binding domain"
FT Binding_site 386..394
FT /note= "Putative cation binding domain"
FT Binding_site 441..449
FT /note= "Putative cation binding domain"
FT Domain 1040..1044
FT /label= Cytoplasmic_domain
FT /note= "Conserved in virtually all integrins"
PN WO9219647-A.
PD 12-NOV-1992.
PF 27-APR-1992; U03527.
PR 03-MAY-1991; US-693564.
PA (SCRI) SCRIPPS RES INST.
PI Quaranta V, Tamura RN;
DR WPI: 92-398799/48.
DR N-PSDB: Q31188.
PT Integrin alpha sub-unit cytoplasmic domain polypeptide(s) - used
PT for prodn. of antibodies and in detection of integrin sub-units
PT in body samples
PS Disclosure; Page 69-73; 115pp; English.
CC The sequences given in R28821-22 are the human alpha 6A and 6B
CC integrin subunits. Integrins are a family of cell surface receptors
CC which serve cellular adhesion functions. These receptors form a link
CC between the extracellular matrix and the cytoskeleton through their
CC binding to various extracellular components. Each integrin receptor
CC is a heterodimer comprised of an alpha and a beta subunit. Each alpha
CC subunit tends to associate with only one type of beta subunit but
CC there are several exceptions to this rule. These integrins correspond
CC to the laminin receptor. The cytoplasmic domain of the 6A and 6B
CC integrins differs from previously isolated alpha 6 integrins. The
CC human alpha 6B was isolated from human choriocarcinoma cell line JAR
CC by immunoprecipitation studies.
SQ Sequence 1073 AA;

Query Match 4.4%; Score 94; DB 5; Length 1073;
Best Local Similarity 25.3%; Pred. NO. 2.21e+01;
Matches 20; Conservative 23; Mismatches 29; Indels 7; Gaps 7;
Db 997 vrvtfpskt-vagysgvpwvllvallagilmallvflwkcggfk-rnk-kdhydat 1053
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Qy 207 VAVSIYGLSRVSDYK-HHWSDLTGLIOGALV-AILVAV-YVSDFFKERTSFKERKEED 263

Db 1054 ykkaeihaqpsdkerltsd 1072
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 264 SHTT-LHETPTTGNHYPSN 281
Search completed: Tue Nov 4 10:20:24 1997
Job time : 94 secs.

This Page Blank (uspto)

WIREA

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Nov 4 10:17:27 1997; MasPar time 10.95 Seconds
Tabular output not generated. 749.603 Million cell updates/sec

Title: >US-08-842-827-2
Description: (1-284) from US08842827.pap
Perfect Score: 2125
Sequence: 1 MFDRKRLPYVALDVLCLLA.....HTTLHETPTGNHPSNHOP 284

Scoring table: PAM 150
Gap 11

Searched: 91006 seqs, 2888923 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir51

1:ann1 2:ann2 3:ann3 4:ann4 5:unann1 6:unann2 7:unann3
8:unann4 9:unann5 10:unann6 11:unann7 12:unann8
13:unann9 14:unann10 15:unenc 16:unrev

Statistics: Mean 45.994; Variance 109.223; scale 0.421

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description	Pred. No.
1	188	8.8	274	11	S69561	hypothetical protein 1.94e-11
2	181	8.5	289	11	S70114	hypothetical protein 1.86e-10
3	121	5.7	345	1	QXKL2M	NADH dehydrogenase (1.26e-02
4	117	5.5	216	10	S32217	hypothetical protein 3.73e-02
5	110	5.2	437	9	H47070	probable transmembra 2.37e-01
6	111	5.2	2220	13	A45290	calcium channel prot 1.83e-01
7	103	4.8	370	9	A44545	amine dehydrogenase 1.41e-00
8	102	4.8	1873	14	A30063	dihydropyridine rece 1.80e+00
9	102	4.8	2208	12	A37860	calcium channel prot 1.80e+00
10	99	4.7	157	13	S58012	probable olfactory r 3.76e+00
11	100	4.7	261	8	I38164	hypothetical protein 2.95e+00
12	99	4.7	342	16	S68129	NADH dehydrogenase (3.76e+00
13	99	4.7	352	5	B34284	NADH dehydrogenase (3.76e+00
14	99	4.7	368	9	S54809	linoleoyl-CoA desatu 3.76e+00
15	99	4.7	382	10	E64209	prolipoprotein diacy 3.76e+00
16	100	4.7	3388	4	GNWVDP	genome polypprotein - 2.95e+00
17	97	4.6	346	5	S10188	NADH dehydrogenase (6.09e+00
18	97	4.6	366	10	B64413	hypothetical protein 6.09e+00
19	98	4.6	676	5	S54750	cyck protein - Rhizo 4.79e+00
20	97	4.6	723	10	D42707	H+/K+-transporting A 6.09e+00
21	97	4.6	885	8	A25817	nonstructural protei 6.09e+00

22	97	4.6	1127	11	S47445	MDM1 protein - yeast 6.09e+00
23	98	4.6	3391	4	GNWV16	genome polypprotein - 4.79e+00
24	98	4.6	3391	4	GNWV26	genome polypprotein - 4.79e+00
25	97	4.6	3391	8	JS0219	polypprotein - dengue 6.09e+00
26	96	4.5	112	16	S61394	hypothetical protein 7.73e+00
27	95	4.5	327	9	I41283	phenylalanine--trNA 9.79e+00
28	95	4.5	331	2	SYECFA	phenylalanine--trNA 9.79e+00
29	95	4.5	346	5	S25423	NADH dehydrogenase (9.79e+00
30	96	4.5	347	5	S41836	NADH dehydrogenase (9.79e+00
31	96	4.5	347	5	S26152	NADH dehydrogenase (7.73e+00
32	96	4.5	397	16	S60334	purinoceptor P2X - r 7.73e+00
33	96	4.5	397	14	I58099	gene P2X3 protein - r 7.73e+00
34	96	4.5	615	14	S43285	noradrenaline transp 7.73e+00
35	95	4.5	727	9	A32561	probable cadmium-tra 9.79e+00
36	94	4.4	157	14	S58031	probable olfactory r 1.24e+01
37	94	4.4	266	5	S54146	trypsin (EC 3.4.21.4 1.24e+01
38	93	4.4	274	10	A26161	halorhodopsin precur 1.56e+01
39	94	4.4	340	4	MMBES	cell fusion protein 1.24e+01
40	94	4.4	427	11	S44549	probable membrane pr 1.24e+01
41	93	4.4	443	5	F30010	NADH dehydrogenase (1.56e+01
42	93	4.4	693	12	S64904	probable membrane pr 1.56e+01
43	94	4.4	948	2	PXMUP2	H+-transporting ATPa 1.24e+01
44	94	4.4	1073	13	B36429	integrin alpha-6 cha 1.24e+01
45	92	4.3	835	8	A27211	virA protein - Agrob 1.97e+01

ALIGNMENTS

RESULT 1
ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#submission
#description
#accession
#molecule-type DNA
#residues
#cross-references EMBL:U33057
GENETICS
#map_position 4R
SUMMARY
Query Match 8.8%; Score 188; DB 11; Length 274;
Best Local Similarity 34.4%; Pred. No. 1.94e-11;
Matches 52; Conservative 33; Mismatches 57; Indels 9; Gaps 9;
Db 107 isndf-hfmhtsilclnliisinaaltgaiklignlpdfrdcipdqkmsdslvf 165
Qy 90 IRNNYATIAYKAGTFLGAAASQSOLDIAKYSIGRLRPHFLDVCDDPSKINSDDGI- 148
Db 166 gldickqtnkwllyeglkstpsghsfivstmgft-yfwgrv-ft-trntsrswppll 221
Qy 149 EYYICRGNAEVRK-EGRLSPYSYSHSFSMYCMLFVALYLAQRMKGDWARLLRPTLQFGLV 207
Db 222 al-v-vmvsvrvidhrhwydvvgavlaflv 250
Qy 208 AVSIYVGLSRVSDYKHHWSDVLTGLIOGALV 238
RESULT 2
ENTRY
TITLE
ORGANISM
DATE
#formal_name Saccharomyces cerevisiae
#sequence_revision 06-Sep-1996 #text_change 06-Sep-1996

```
ACCESSIONS S70114
REFERENCE S70114
#authors Fulton, L.
#submission submitted to the EMBL Data Library, May 1996
#description The sequence of S. cerevisiae cosmid 9819.
#accession S70114
#molecule_type DNA
#residues 1-289 ##label FUL
#cross-references EMBL:U51031
GENETICS
#map_position 4R
SUMMARY #length 289 #molecular-weight 33514 #checksum 4703
Query Match 8.5%; Score 181; DB 11; Length 289;
Best Local Similarity 27.0%; Pred. No. 1.86e-10;
Matches 67; Conservative 65; Mismatches 98; Indels 18; Gaps 16;
Db 38 qqpfersqfyndltishpy-atte-rvnnmlfvsfvpstl-llil-gsll-adr-- 89
QY 31 HTPFQGVCFNDESIRYPYKEDTIPYALLGGIIPESIIILGETLSVVCNLLHNSFI 90
Db 90 rh-lifilvtslglawfstffnfknwgrlprdfldrcqpy-eglpdltlftak 147
QY 91 RNNYIATIKATCTFUGAASOSLTDIAKYSIGRUPHFLDVCDDPWSKINGSDGYIEY 150
Db 148 dvcttknherlldgfttp-sghssesfaglyvfwlcgqlitesplmplwrkmvafip 206
QY 151 YICRG-NAERVKEG-RLSFYSGHSSFSMYCMLFVALYLQARKGDWARL-L-RPTLOF-G 205
Db 207 llgaal-lalstrtdyrhfvdlvsgmlgyimahfyrirfpiddplpfkplmd-dsd 264
QY 206 LVAVSIYGLSRVSDYKHHSVDLTGLIOGALVAILVAVVVSDFFKERTSFKERKEEDSH 265
Db 265 vtleeaavt 272
QY 266 TLUHEPT 273
RESULT 3
ENTRY QXXL2M #type complete
TITLE NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 -
ALTERNATE_NAMES African clawed frog mitochondrion (SCC1)
ORGANISM NADH-ubiquinone oxidoreductase chain 2
#formal_name Mitochondrion Xenopus laevis #common_name
#accession African clawed frog
DATE 28-Feb-1986 #sequence_revision 28-Feb-1986 #text_change
ACCESSIONS A00417
REFERENCE A00155
#authors Roe, B.A.; Ma, D.P.; Wilson, R.K.; Wong, J.F.H.
#journal J. Biol. Chem. (1985) 260:9759-9774
#title The complete nucleotide sequence of the Xenopus laevis
#cross-references M01D:85261388
#accession A00417
#molecule_type DNA
#residues 1-345 ##label ROE
GENETICS
#genome mitochondrion
#genetic_code SCC1
CLASSIFICATION #superfamily NADH dehydrogenase (ubiquinone) chain 2
KEYWORDS membrane-associated complex; mitochondrion; NAD; oxidative
phosphorylation; oxidoreductase; respiratory chain
SUMMARY #length 345 #molecular-weight 37671 #checksum 3982
Query Match 5.7%; Score 121; DB 1; Length 345;
Best Local Similarity 25.3%; Pred. No. 1.26e-02;
Matches 21; Conservative 27; Mismatches 28; Indels 7; Gaps 7;
Db 51 raleastkvfltaqaasal-llfssin-nawltgwsildtlnplscatmtaiaimkfg- 107
QY 159 RVKEGRLSFYSGHSSFSMYCMLFVALYLQARKGDWARL-LRPTLOGLVAVSIYGLSR 217
```

```
Db 108 lapf-hfwlpevlqglstttgli 129
QY 218 VSDYKHHW-SDVLTGL-IQGALV 238
RESULT 4
ENTRY #type complete
TITLE hypothetical protein 2 - Bacillus megaterium
ORGANISM #formal_name Bacillus megaterium
DATE 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change
ACCESSIONS S32217
REFERENCE S32215
#authors Rauschenbach, R.; Isernhaagen, M.; Noeske-Jungblut, C.;
#submission Boldol, W.; Siewert, G.
#description submitted to the EMBL Data Library, March 1993
#description Cloning, sequencing and expression of the gene for cytochrome
P450meg, the steroid 15beta-monooxygenase from Bacillus
megaterium ATCC 13368.
#accession S32217
#status preliminary
#molecule_type DNA
#residues 1-216 ##label RAU
#cross-references EMBL:Z21972
SUMMARY #length 216 #molecular-weight 24946 #checksum 8213
Query Match 5.5%; Score 117; DB 10; Length 216;
Best Local Similarity 29.4%; Pred. No. 3.73e-02;
Matches 32; Conservative 31; Mismatches 37; Indels 9; Gaps 8;
Db 114 rpdhlrlldigystfsgpshannafslgyl-tfl-lwrhitarwaril-lll-fsmlmi- 168
QY 154 RGNARVKE-GRLSFYSGHS--SFSMYCMLFVALYLQARKGDWARLRLPTLOGLVAVS 210
Db 169 lsiglsrlygvhypsdiagylagg-cwiaislwfgyqdrknkdr 216
QY 211 IYVGLSRVSDYKHHWSVDLTGLIOGALVAILVAVVVSDFFKERTSFKER 259
RESULT 5
ENTRY H47070 #type complete
TITLE probable transmembrane export protein - Yersinia
ORGANISM pseudotuberculosis
DATE 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change
ACCESSIONS H47070
REFERENCE A47070
#authors Kessler, A.C.; Haase, A.; Reeves, P.R.
#journal J. Bacteriol. (1993) 175:1412-1422
#title Molecular analysis of the 3,6-dideoxyhexose pathway genes of
Yersinia pseudotuberculosis serogroup IIA.
#cross-references M01D:93186709
#accession H47070
#status preliminary
#molecule_type nucleic acid
#residues 1-437 ##label KES
#cross-references NCBI:126815; NCBI:126836
#experimental_source serogroup IIA, M85
#note sequence extracted from NCBI backbone
SUMMARY #length 437 #molecular-weight 49429 #checksum 4248
Query Match 5.2%; Score 110; DB 9; Length 437;
Best Local Similarity 37.5%; Pred. No. 2.37e-01;
Matches 21; Conservative 16; Mismatches 16; Indels 3; Gaps 3;
Db 89 iipifilfilltltspyselggfllnngqrnsnifk-vasfitttsignla 143
QY 63 IIPFSIIIVILLGETLSVY-CN-LLHNSNFIRNNYIATYKAIGTFLFGAASQSULT 116
RESULT 6
ENTRY A45290 #type complete
```


[illegible]

[illegible]

This Page Blank (uspto)

[illegible]

KW HYDROLASE; HYDROGEN ION TRANSPORT; TRANSMEMBRANE; PHOSPHORYLATION;
KW ATP-BINDING; MULTIGENE FAMILY.
FT NON_TER 1 1
FT TRANSMEM <1 15 3 (POTENTIAL).
FT DOMAIN 16 27 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 28 50 4 (POTENTIAL).
FT DOMAIN 51 425 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 426 444 5 (POTENTIAL).
FT DOMAIN 445 460 6 (POTENTIAL).
FT TRANSMEM 461 484 7 (POTENTIAL).
FT DOMAIN 485 542 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 543 563 8 (POTENTIAL).
FT DOMAIN 562 578 9 (POTENTIAL).
FT TRANSMEM 579 595 10 (POTENTIAL).
FT DOMAIN 586 704 11 (POTENTIAL).
FT MOD_RES 81 81 12 (POTENTIAL).
FT BINDING 175 175 13 (POTENTIAL).
SQ SEQUENCE 704 AA; 78039 MW; C1CB2D8A CRC32;

Query Match 4.6%; Score 98; DB 7; Length 704;
Best Local Similarity 27.4%; Pred. NO. 1.02e+00;
Matches 23; Conservative 19; Mismatches 38; Indels 4; Gaps 4;

Db 480 fwaayet-qfprvfgvstlqrtatddfrklasaikyqvstisqalifvtrswsfver 538
Qy 167 FYSGHSSFSMYCMLFVALYLOARMKGDWARLLRPT-LQFGLVAVS-IYVGLSRVSDYKHH 224
Db 539 pglilvvalivavqlvatliavvas 562
Qy 225 WSDVLT-GLIOGALVAILVAVVS 247

Search completed: Tue Nov 4 10:17:09 1997
Job time : 47 secs.

This Page Blank (uspto)

M P S R E L

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPSrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Thu Nov 6 10:42:02 1997; Maspar time 1306.58 seconds
Tabular output not generated. 1316.341 Million cell updates/sec

Title: >US-08-842-827-3
Description: (1-1566) from US08842827.seq
Perfect Score: 1566
N.A. Sequence: 1 CCTGTGGGAGAGCGCGG.....CCAAAAA.....GGTTTTTTTTTTTT
Comp: GGACACCTCTCTCGCGCC.....

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0: Query 0

Searched: 362067 seqs, 549138275 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb1-new3
1-BCT 2:FUN 3:GEN1 4:GEN2 5:HTG1 6:HTG2 7:HUM 8:INV
9:ORG 10:MAM 11:VRT 12:PLN 13:PRO 14:ROD 15:SYN 16:UNC
17:VIR

Database: genbank99
18:BCT1 19:BCT2 20:BCT3 21:BCT4 22:BCT5 23:BCT6 24:BCT7
25:BCT8 26:BCT9 27:BCT10 28:BCT11 29:GEN1 30:GEN2
31:GEN3 32:HTG1 33:HTG2 34:HTG3 35:INV1 36:INV2 37:INV3
38:INV4 39:INV5 40:INV6 41:INV7 42:INV8 43:INV9 44:INV10
45:INV11 46:MAM1 47:MAM2 48:MAM3 49:VRT1 50:VRT2 51:VRT3
52:VRT4 53:PAT1 54:PAT2 55:PAT3 56:PAT4 57:PAT5 58:PHG
59:PLN1 60:PLN2 61:PLN3 62:PLN4 63:PLN5 64:PLN6 65:PLN7
66:PLN8 67:PLN9 68:PLN10 69:PLN11 70:PLN12 71:PLN13
72:PLN14 73:PLN15 74:PLN16 75:PLN17 76:PLN18 77:PLN19 78:PLN20
79:PLN21 80:PLN22 81:PLN23 82:PLN24 83:PLN25 84:PLN26
85:ROD1 86:ROD2 87:ROD3 88:ROD4 89:ROD5 90:ROD6 91:ROD7
92:ROD8 93:STR 94:SYN 95:UNA 96:VRL1 97:VRL2 98:VRL3
99:VRL4 100:VRL5 101:VRL6 102:VRL7 103:VRL8 104:VRL9
105:VRL10

Database: genbank-new3
106:BCT 107:GEN1 108:GEN2 109:HTG1 110:HTG2 111:INV
112:MAM 113:VRT 114:PHG 115:PLN 116:PRI1 117:PRI2
118:ROD 119:SYN 120:UNA 121:VRL

Database: u-emb150_99
122:part1

Statistics: Mean 11.944; Variance 6.199; scale 1.927

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.	
1	429	27.4	1212	85	D84376	House mouse; Musculus	0.00e+00	
2	265	16.9	1490	88	MUSHP1P	Mus musculus (clone H	4.61e-178	
c	130	8.3	3353	82	HUMORFA02	Human mRNA for KIAA00	7.06e-72	
4	111	7.1	2206	91	RNDRI42	R.norvegicus mRNA for	1.76e-57	
5	102	6.5	1444	78	HSU79294	Human clone 23748 mRN	9.39e-51	
c	41	2.6	215	57	128278	Sequence 5 from paten	2.48e-08	
7	34	2.2	215	57	128278	Sequence 5 from paten	3.13e-04	
8	32	2.0	1727	45	SSM2RR	S.solidissima mRNA fo	4.01e-03	
c	9	31	2.0	4601	40	DMU11584	Drosophila melanogast	1.39e-02
c	10	31	2.0	14626	35	BNMFCEG	Bombyx mori fibroin l	1.39e-02
c	11	31	2.0	19517	40	DMU37541	Drosophila melanogast	1.39e-02
12	30	1.9	1427	39	DDAAC3	Dictyostelium discoid	4.73e-02	
13	29	1.9	1512	46	BRCA	B.taurus mRNA for cyc	1.57e-01	
14	29	1.9	1837	91	RNP2XMR	R.norvegicus P2X mRNA	1.57e-01	
15	30	1.9	1862	64	PSMANC	Pyromyces sp. mRNA fo	4.73e-02	
16	29	1.9	8567	16	CAGNAT2	C.aureus GNAT2 gene	1.57e-01	
c	17	28	1.8	363	91	RNA7CINT	R.norvegicus mRNA for	5.09e-01
18	28	1.8	417	40	DOATRICH	D.obscura A+T-rich re	5.09e-01	
c	19	28	1.8	793	122	MMD430	Mouse Mur1 mRNA, exo	5.09e-01
20	28	1.8	793	118	D85430	Mouse Mur1 mRNA, exo	5.09e-01	
21	28	1.8	793	14	MMD430	Mouse Mur1 mRNA, exo	5.09e-01	
22	28	1.8	1524	43	LPACT3	L.polyphemus mRNA for	5.09e-01	
23	28	1.8	1737	39	DD19	Dictyostelium discoid	5.09e-01	
24	28	1.8	2049	44	S67872	Zw-glucose-6-phosphat	5.09e-01	
c	25	28	1.8	2630	43	PCU53325	Plasmodium chabaudi d	5.09e-01
26	28	1.8	3077	44	PFU07706	Plasmodium falciparum	5.09e-01	
27	28	1.8	3774	91	RNIAC	R.norvegicus mRNA for	5.09e-01	
28	28	1.8	6372	39	DDICNPA	Dictyostelium discoid	5.09e-01	
c	30	28	1.8	10359	44	PFARPI	Plasmodium falciparum	5.09e-01
c	31	28	1.8	192274	110	HS435C23	Human DNA sequence **	5.09e-01
32	27	1.7	347	54	A23327	O.sativa mRNA for T23	5.09e-01	
33	27	1.7	451	60	BNU59459	Brassica napus antiu	1.61e+00	
34	27	1.7	825	65	SCCOX1	Yeast cytochrome c ox	1.61e+00	
c	35	27	1.7	1095	61	DDICSA	Dictyostelium discoid	1.61e+00
36	27	1.7	1128	43	MPU15677	Myrmecia pilosula H18	1.61e+00	
37	27	1.7	1381	39	CHTRP45MR	C.tentans mRNA for hr	1.61e+00	
38	27	1.7	1702	43	PFAGPTA	Plasmodium falciparum	1.61e+00	
39	27	1.7	1803	73	HSARGBPTA	H.sapiens mRNA for Ar	1.61e+00	
c	40	27	1.7	2035	39	DDU53884	Dictyostelium discoid	1.61e+00
41	27	1.7	2045	44	PNLGTSYN	Parulirus argus gluta	1.61e+00	
42	27	1.7	2798	42	HYDHP701B	Hydra magnipapillata	1.61e+00	
c	43	27	1.7	3700	39	DDGP80G	D.discoidium gp80 gen	1.61e+00
44	27	1.7	7989	44	PFU31083	Plasmodium falciparum	1.61e+00	
45	27	1.7	9353	44	PFU67959	Plasmodium falciparum	1.61e+00	
ALIGNMENTS								
RESULT	1	D84376	1212 bp	mRNA	03-OCT-1996	ROD		
LOCUS								
DEFINITION		House mouse; Musculus domesticus kidney mRNA for Phosphatidic acid phosphatase, complete cds.						
ACCESSION		D84376						
NID		91487872						
KEYWORDS		Phosphatidic acid phosphatase; 35-kDa phosphatidic acid phosphatase.						
SOURCE		Mus musculus						
ORGANISM		Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.						
REFERENCE		1 (bases 1 to 1212)						
AUTHORS		Kanoh, H.						
TITLE		Direct Submission						
JOURNAL		Submitted (10-APR-1996) to the DDBJ/EMBL/GenBank databases. Hideo Kanoh, Sapporo Medical University School of Medicine, Department of Biochemistry, West-17, South-1, Sapporo, Hokkaido 060, Japan (E-mail:kanoh@serpent.cc.sapmed.ac.jp, Tel:011-611-2111(ex.2290), Fax:011-612-5861)						

REFERENCE 2 (bases 1 to 1212)
AUTHORS Kai.M., Wada,I., Imai,S., Sakane,F. and Kanoh,H.
TITLE Identification and cDNA cloning of 35-kDa phosphatidic acid phosphatase (type 2) bound to plasma membranes. Polymerase chain reaction amplification of mouse H2O2-inducible hic53 clone yielded the cDNA encoding phosphatidic acid phosphatase
JOURNAL J. Biol. Chem. 271 (31), 18931-18938 (1996)
MEDLINE 96324980
FEATURES
Source Location/Qualifiers
1..1212
/organism="Mus musculus"
/tissue_type="kidney"
287..1138
/evidence=experimental
/codon_start=1
/product="Phosphatidic acid phosphatase"
/db_xref="PID:gl487873"
/translation="MFDTRLPYVALDVICVLGAPFAILTSRHTPFGIFCNDSDS IKPYKEDTIPYALGIVIPFCIIIVMSIGESLSYFVNLHNSVFNPIYATIKAV GAFLEVSASQSLTDIAKVTIGSLRPHFAICNPWSKINCSDGYEDYICOGNEKV KEGRLSFYSGHSSFMYCMLFVALYLOAKMKDWARLLRPMLOFGLIAFSIYVGLSRV SDYKHMSDVTYVGLLOGAMALVALYVSDFFEDTHSKERKEEDPHILHETASRN YSTNEP"
BASE COUNT 277 a 316 c 313 g 306 t
ORIGIN
Query Match 27.4%; Score 429; DB 85; Length 1212;
Best Local Similarity 76.8%; Pred. No. 0.00e+00;
Matches 636; Conservative 0; Mismatches 189; Indels 3; Gaps 2;
Db 274 ccagccggtgaccatgttcgacagacgagcggtcgccgtacgtgcccctcgatgtattg 333
QY 329 CCGGCGAGAGACCATGTTGACAAGACGGCGGTGCGGTAGTGGCCCTCGATGCTCTG 368
Db 334 cgtgtgtcgtggtgattgaccttttgcattcttcaattcttactcaaggca--taccoccc-ttcca 390
QY 389 CGFTGTTGCTGGCTTCCATGCTATGGCTGTCTTAAATTTGGGCAATATATCATTTCA 448
Db 391 gcgaggaattctgtaagtacactccatcaagtacccttacaaggaagacacatacc 450
QY 449 GAGAGGCTTTTCTGTAAAGACAACAGCATCAACTATCCGTACATGACAGTACCCGCCG 508
Db 451 ttatgccttattagggtggaatagtcattccattcctctgtattctgtattgattggaga 510
QY 509 ATCCACTGTCCTCATCTAGTGGGGTGGCTTGCCGGTTCCTCTATATTCTTGGAGA 568
Db 511 atctcgtctgttacttaagtctctgattcgaattcctttgtggaacatccctacat 570
QY 569 AACCCGTCTGTGTACTGTAAACCTTTTGCACTCAAAATTCCTTTATCAGTAATAACATAC 628
Db 571 agccaccatttacaagccgctcgagccctttttgttcgagtcctcagctcagtcctt 630
QY 629 AGCCATATTTCAGAACCCATTTGGAACCTTTTATTGTTGGTGCAGCTAGTCAAGTCCCT 688
Db 631 gactgacatcgctaaagtatactactataggcaggtttcgccgcgacacttcttggctatctgtaa 690
QY 689 GACTGACATGCGCACTATTAATAGCAGACTGCGGCCCTCACTCTTGGATGTTGTGA 748
Db 691 ccagactcgttcaaaaatacaactcgsgtgcattgactatattgaggactacatatgtcaagg 750
QY 749 TCCAGATTGCTGTAATAAATCAACTGCAGCATGTTTACATTGAATACATACATATGTCGAG 808
Db 751 gaatgaagaaagtcgaaggagcgaggttctcttctactcgggacactcttcttc 810
QY 809 GAATGCAGAAAGAGTTAAGGAAGGAGGTGTCCTTCTATTACAGGCCACTCTTTCGTTTTC 868
Db 811 tatgtactcgtcgtgtttgtcgcacatttatcttcaagccaggtatgaaggagactgggccc 870
QY 869 CATGTACTGCATCTGTTGTGGCACTTTATCTTCAAGCCAGGATGAGGGAGACTGGGC 928
Db 871 aagactcttacgaccatcgtccagtttggggtcattgtcttttccatatatgtgggacct 930
QY 929 AAGACTCTTACGCCCCACACTGCAATTTGGTCTGTTGGCGTATTCATTTATGTGGGCT 988

Db 931 ttctcagtgctctactacaaacacacactggagtgacgtccacagttggactcattcaggg 990
QY 989 TTCTCGAGTTCCTGATTATAAACACACACTGGAGCGATGTTGACTGGACTCATTCAGGG 1048
Db 991 agctcgtatggtctactctgttctgttatgtatccgattcttccaaggacacacattc 1050
QY 1049 AGCTCTGTTGCAATATTAGTTGCTGTATATGTATCGGATTTCTTCAAGAAGAAGACTTC 1108
Db 1051 ttacaaagagagaaggaagagatccacacacagctctccatgaaac 1098
QY 1109 TTTTAAAGAAAGAAAGAGGAGGACTCTCATACAACCTGTCATGAAC 1156
RESULT 2
LOCUS MUSHPIP 1490 bp mRNA ROD 11-JAN-1996
DEFINITION Mus musculus (clone HIC-53) hydrogen peroxide-inducible protein mRNA, complete cds.
ACCESSION L43371
NID g1161099
KEYWORDS hydrogen peroxide; hydrogen peroxide-inducible protein.
SOURCE Mus musculus (clone: HIC-53) embryo calvariagag cDNA to mRNA.
ORGANISM Mus musculus
Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae; Murinae; Mus
REFERENCE 1 (bases 1 to 1490)
AUTHORS Egawa,K., Yoshiwara,M., Shibamura,M. and Nose,K.
TITLE Isolation of a novel ras-reversion gene that is induced by hydrogen peroxide from a mouse osteoblastic cell line, MC3T3-E1
JOURNAL FEBS Lett. 372 (1), 74-77 (1995)
MEDLINE 96032549
FEATURES Location/Qualifiers
source 1..1490
/organism="Mus musculus"
/clone="HIC-53"
/cell_line="MC3T3-E1"
/cell_type="osteoblast"
/dev_stage="embryo"
/tissue_type="calvariagag"
<1..978
/codon_start=1
/product="hydrogen peroxide-inducible protein"
/db_xref="PID:gl161100"
/translation="ARASLELRPVGGTGORRIGRDEPALRQAPQAIASTVVK ESRKHSVGRPIQTGPQPCSRQKRLAVAGRRHPEGALGIGYLDRRGLPLPAP GGDITQPVTFMEDKTRLPYVALDVICVLGAPFAILTSRHTPFGIFCNDSDIKPY KEDTIPYALGIVIPFCIIIVMSIGESLSYFVNLHNSVFNPIYATIKAVGAFLE GVSASQSLTDIAKVTIGSLRPHFAICNPWSKINCSDGYEDYICOGNEKVKEGRL SFSTGLFTLYVLHAVCRTLSRRRLARLLRPMLOFGLIAFSIYVGLFSSV"
CDS
1..1490
3'UTR 979..1490
BASE COUNT 362 a 367 c 368 g 393 t
ORIGIN
Query Match 16.9%; Score 265; DB 88; Length 1490;
Best Local Similarity 75.1%; Pred. No. 4.61e-178;
Matches 434; Conservative 0; Mismatches 139; Indels 5; Gaps 3;
Db 321 ccagccggtgaccatgttcgacacagcggctcgctacgtgcccctcgatgtattg 380
QY 329 CCGGCGAGAGACCATGTTTGACAAGACGGCGTGGCTAGTGGCCCTCGATGCTCTG 388
Db 381 cgtgtgtcgtggtgattgaccttttgcattcttacttcaaggca--taccoccc-ttcca 437
QY 389 CGTGTGCTGGCTTCCATGCCCTATGCTGTCTTAAATTTGGCCAAATATATCATTTCA 448
Db 438 gcgaggaatattctgaatgacatccatcagcttacccttacaaggaagacacatacc 497
QY 449 GAGAGCTTTTCTGTAAAGACAACAGCATCAACTATCCGTACCATGACAGTACCCGCCG 508
Db 498 ttatgccttattagggtggaatagtcattccattctgtattctcgttattgattggaga 557

[illegible]


```

RESULT 13
LOCUS BCCA 1512 bp RNA MAM 19-JUL-1996
DEFINITION B. taurus mRNA for cyclin A.
ACCESSION X68321
NID 99
KEYWORDS cyclin A; protein kinase activation.
SOURCE domestic cattle or domestic cow.
ORGANISM Bos taurus
REFERENCE 1 (bases 1 to 1512)
AUTHORS Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.
Hunt,T.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-1992) T. Hunt, ICRF Clare Hall Laboratories,
South Mimms, Herts EN6 3LD, UK
REFERENCE 2 (bases 1 to 1512)
AUTHORS Kobayashi,H., Stewart,E., Poon,R., Adamczewski,J.P., Gannon,J. and
Hunt,T.
TITLE Identification of the domains in cyclin A required for binding to,
and activation of, p34cdc2 and p32cdk2 protein kinase subunits
Mol. Biol. Cell 3 (11), 1279-1294 (1992)
JOURNAL 93091274
REFERENCE 3 (bases 433 to 1221)
AUTHORS Brown,N.R., Noble,M.E., Endicott,J.A., Garman,E.F., Wakatsuki,S.,
Mitchell,E., Rasmussen,B., Hunt,T. and Johnson,L.N.
TITLE The crystal structure of cyclin A
JOURNAL Structure 3 (11), 1235-1247 (1995)
MEDLINE 9616440
COMMENT In the 'Structure' publication: The 5' end of clone A3 was made by
PCR, and introduced a methionine in place of a valine, followed by
a glycine to make an NcoI site. At the 3' end, an XhoI site was
engineered in to allow subcloning into pET21d, which also puts 6
histidines at the C-terminus that were used for purification.
FEATURES
source
1..1512
/organism="Bos taurus"
/dev_stage="adult"
/tissue_type="lymphocyte"
<1..1221
/notes="subunit structure-cyclin A /p34cdc2 or cyclin A
/p32cdk2"
/codon_start=1
/product="cyclin A-3"
/db_xref="PID:e54213"
/db_xref="PID:g10"
/db_xref="SWISS-PROT:P30274"
/translation="EFQEDQENVNPEKAAPQAPQPRTRAGLAVLRAGNSRGAPQRPKT
RRVAPLKDLPINDEYVPPWPKANNKQAPFTIHVDAEETQKRPTEKSKSESDYLAF
NSAVTLPGPKPLAPLDYPMDSFSPHTMEMSVLDEKPVSNVEVPDYHEDIYTL
REMEVKCKPKVGMKKQPDITNSMRALVDWLVEGVEYKLQNETLHAVNYIDRFLS
MSVLGRKLQLVGTAAMLLASKFEEIYPPPEAEFVITDDTYTKQVLRHEHLVRLVL
AFDLAAPTINQFTQYFLHQOPANCKVESLAMEFLGSLIDADPYLKLPLSPVITAAAF
HLALYVYGQSWPESLVQKTYGTLETLPCLLDLHQTLYLRAPQHAQOOSIREKYKNSY
HGVSLLNPPETLVN"
BASE COUNT 481 a 298 c 321 g 412 t
ORIGIN
Query Match 1.9%; Score 29; DB 46; Length 1512;
Best Local Similarity 72.3%; Pred. No. 1.57e-01;
Matches 47; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
Db 1448 tattaatgatgcagtcgagcataataataatgattgattgataaaaaa 1507
|||||
QY 1496 TATGAATGATGTTGATTAAATATATACATATATAAATGATGGAGACCAAAAA 1555
Db 1508 aaaa 1512
|||||
QY 1556 AAAAA 1560
RESULT 14
LOCUS RNP2XMR 1837 bp RNA ROD 06-JAN-1995

```

```

DEFINITION R. norvegicus P2X mRNA.
ACCESSION X80477
NID 9558239
KEYWORDS ATP receptor; P2X gene.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
REFERENCE 1 (bases 1 to 1837)
AUTHORS Valera,S., Hussy,N., Evans,R.J., Adami,N., North,R.A.,
Surprenant,A. and Buell,G.
TITLE A new class of ligand-gated ion channel defined by P2x receptor for
extracellular ATP
JOURNAL Nature 371 (6497), 516-519 (1994)
MEDLINE 95021681
REFERENCE 2 (bases 1 to 1837)
AUTHORS Buell,G.N.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-1994) G.N. Buell, Glaxo Inst of Mol Biology, 14
Chemin des Aulx, 1228 Plan-les-Ouates, Geneva, SWITZERLAND
FEATURES
source
1..1837
/organism="Rattus norvegicus"
/cell_type="smooth muscle"
/clone_lib="pBRCMV plasmid cDNA library"
210..1409
/gene="P2X"
/codon_start=1
/db_xref="PID:g558240"
/translation="MARLQDELSAFFFEYDTPRMVLVRNKKVGIVRILQVLVLYV
IGWVFYERQYOTSSLVSVKLGAVTQGLGQVQVADVTFPAHGDSSFV
MTNFVTPQOTQCHCAENPEGGICODDSCGTPGAERKAQGTGCVFNGTVKTC
ITGWCPEVDDKIPSPALLIREAENTLFIKNSISPRFKVNRNLVEEVNGYMKKCL
YHKIQHPLCPVNLGYVVRSCQDPSRLAEKGGVGTIDWKCDLDDHVRHCKPIYQF
HGLYGEKNLSPGFNRFAHFVQNGTNRHLEKVFGEHFDILVDGRAGAFDILPTMT
IGSGIGFGVATVLCDLLLHLPLKRRHYKQKFKYAEDMGPEGEHDPVATSTLGL
QENMRYS"
BASE COUNT 444 a 477 c 472 g 444 t
ORIGIN
Query Match 1.9%; Score 29; DB 91; Length 1837;
Best Local Similarity 72.3%; Pred. No. 1.57e-01;
Matches 47; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
Db 1771 tgaacttaattaaataataataaaaaa 1830
|||||
QY 1502 TGATGTTGATTAAATATATACATATATAAATGATGGAGACCAAAAAA 1561
Db 1831 aaaa 1835
|||||
QY 1562 AAAAA 1566
RESULT 15
LOCUS PSMANC 1862 bp RNA PLN 07-FEB-1997
DEFINITION Pimomyces sp. mRNA for endo-b1,4-mannanase.
ACCESSION X97520
NID 91292878
KEYWORDS endo-1,4 beta-mannanase; manC gene.
SOURCE Pimomyces sp.
ORGANISM Pimomyces
REFERENCE 1 (bases 1 to 1862)
AUTHORS Gilbert,H.J.
TITLE Direct Submission
JOURNAL Submitted (24-APR-1996) H.J. Gilbert, Department of Biological and
Nutritional Sciences, Univ. of Newcastle upon Tyne, Newcastle, NE1
7RU, UK
REFERENCE 2 (bases 1 to 1862)
AUTHORS Millward-Sadler,S.J., Hall,J., Black,G.W., Hazlewood,G.P. and

```

W P S R E H
(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.
MPSrch_n n.a. n.a. database search, using Smith-Waterman algorithm
Run on: Thu Nov 6 11:06:47 1997; Maspar time 169.89 Seconds
Tabular output not generated.
Title: >US-08-842-827-3
Description: (1-1566) from US08842827.seq
Perfect Score: 1566
N.A. Sequence: 1 CCTGTGGGAGAGCGCGG.....CCAAAAA.....GGTTTTTTTTTTTT
Comp: GGACACCTCTCTCGCGCC.....GGTTTTTTTTTTTTTTTT
Scoring table: TABLE default
Gap 6
Nmatch STD : Dbase 0; Query 0
Searched: 142080 seqs, 52183452 bases x 2
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: n-geneseq28
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29
Statistics: Mean 9.622; Variance 6.597; scale 1.459
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES							
Result	No.	Score	Match	Length	DB	ID	Pred. No.
C	1	90	5.7	1047	2	Q10572	1.84e-34
C	2	71	4.5	1047	2	Q10572	9.34e-24
C	3	45	2.9	91	9	Q51746	7.85e-10
C	4	41	2.6	91	9	Q51746	8.16e-08
C	5	40	2.6	204	1	N81164	2.56e-07
C	6	38	2.4	204	1	N81164	2.48e-06
C	7	36	2.3	114	12	Q70468	2.28e-05
C	8	36	2.3	114	12	Q70468	2.28e-05
C	9	34	2.2	67	24	T14322	2.04e-04
C	10	35	2.2	114	12	Q70469	6.86e-05
C	11	34	2.2	114	12	Q70465	2.04e-04
C	12	34	2.2	114	12	Q70467	2.04e-04
C	13	35	2.2	114	12	Q70465	6.86e-05
C	14	34	2.2	114	12	Q70470	2.04e-04
C	15	34	2.2	114	12	Q70468	2.04e-04

C	16	34	2.2	114	12	Q70469	Generic DNA sequence	2.04e-04
C	17	33	2.1	138	7	Q42784	Ligand-induced gens,	6.01e-04
C	18	32	2.0	114	12	Q70466	Generic DNA sequence	1.75e-03
C	19	31	2.0	114	12	Q70470	Generic DNA sequence	5.03e-03
C	20	32	2.0	114	12	Q70466	Generic DNA sequence	1.75e-03
C	21	31	2.0	3521	7	Q41226	Clone GP3 encoding ma	5.03e-03
C	22	29	1.9	114	12	Q70472	Generic DNA sequence	4.00e-02
C	23	30	1.9	114	12	Q70472	Generic DNA sequence	1.43e-02
C	24	29	1.9	114	12	Q70473	Generic DNA sequence	4.00e-02
C	25	29	1.9	1837	23	T33852	Rat vas deferens P7x	4.00e-02
C	26	28	1.8	114	12	Q70471	Generic DNA sequence	1.11e-01
C	27	28	1.8	114	12	Q70471	Generic DNA sequence	1.11e-01
C	28	28	1.8	565	6	Q35072	RCV envelope region n	1.11e-01
C	29	28	1.8	1174	9	Q51020	Human FcER1 alpha gen	1.11e-01
C	30	26	1.7	39	7	Q51787	Mixed oligonucleotide	8.04e-01
C	31	26	1.7	74	21	T13613	DC43 TSAR library gen	8.04e-01
C	32	26	1.7	81	21	T13611	DC43 TSAR library gen	8.04e-01
C	33	27	1.7	347	4	Q27482	T23 CDNA.	3.01e-01
C	34	27	1.7	498	3	N50034	Sequence encoding new	3.01e-01
C	35	27	1.7	501	3	N50024	Sequence encoding new	3.01e-01
C	36	27	1.7	501	3	N50023	Sequence encoding new	3.01e-01
C	37	26	1.7	501	3	N50033	Sequence encoding new	8.04e-01
C	38	27	1.7	3520	11	Q69215	Interleukin-10 recept	3.01e-01
C	39	26	1.7	4193	14	Q82834	Human stromalin-2 DNA	8.04e-01
C	40	25	1.6	36	2	Q11195	Ballast Constituent c	2.11e-00
C	41	25	1.6	501	3	N50026	Sequence encoding new	2.11e-00
C	42	25	1.6	961	2	Q10791	Bovine vascular endot	2.11e-00
C	43	25	1.6	961	7	Q44259	Bovine VEGF-164 codin	2.11e-00
C	44	25	1.6	1052	1	N90224	Malaria-specific DNA	2.11e-00
C	45	25	1.6	1599	1	N90733	Genomic DNA from Plas	2.11e-00

ALIGNMENTS

RESULT 1
ID Q10572 standard; DNA; 1047 BP.
AC Q10572;
DT 09-APR-1991 (first entry)
DE Human Natriuretic Peptide Receptor B.
KW NPRB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
hyperaldosteronism; glaucoma; guanylyl cyclase.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= signal sequence
FT Protein 12
FT /label= mature NPRB
FT Domain 23..455
FT /label= extracellular domain
FT /note= "binds natriuretic peptides A,B and C]"
FT Domain 456..456
FT /label= transmembrane domain
FT Domain 479..1047
FT /label= cytoplasmic domain
FT /note= "GC and protien kinase activity"
FT Modified-site 24..26
FT /label= N-glycos_site
FT Modified-site 35..37
FT /label= N-glycos_site
FT Modified-site 161..163
FT /label= N-glycos_site
FT Modified-site 195..197
FT /label= N-glycos_site
FT Modified-site 244..246
FT /label= N-glycos_site
FT Modified-site 277..279
FT /label= N-glycos_site
FT Modified-site 349..351
FT /label= N-glycos_site
FT Modified-site 500..602
FT /label= N-glycos_site
PN WO9100292-A.
PD 10-JAN-1991.

ID Q70469 standard; DNA; 114 BP.
 AC Q70469;
 DT 07-APR-1995 (first entry)
 DE Generic DNA sequence to generate a random TSAR peptide library.
 KW TSAR: totally synthetic affinity reagent; synthetic; binding domain;
 KW effector domain; concatenated heterofunctional protein; linker;
 KW direct; rapid; detection; screening; treatment; generic; ss.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT misc_feature 55..60
 FT /*tag= a
 FT /note= "this sequence represents 'Z'; Z can be a
 FT sequence of 6, 9 or 12 nucleotides (see
 FT comments)"
 PN WO9418318-A.
 PD 18-AUG-1994.
 PF 01-FEB-1994; U00977.
 PR 01-FEB-1993; US-013416.
 PR 30-DEC-1993; US-176500.
 PR 31-JAN-1994; US-189331.
 PA (UYN-) UNIV NORTH CAROLINA.
 PI Fowlkes DM, Kay BK;
 DR WPI: 94-279739/34.
 PT Identifying proteins or peptide(s) which bind a ligand - by
 PT screening a recombinant vector library expressing fusion proteins
 PT comprising a binding domain and an effector domain
 PS Disclosure; Page 35; 255pp; English.
 CC This generic formula can be represented as follows: X(TGC)(NNB)10-
 CC (TGC)(NNB)6Z(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking restriction
 CC sites (X is not the same as Y) that are not specified further. This
 CC sequence generates peptides that are cloverleaf in structure. Other
 CC generic sequences are shown in Q70465-68. Other specific peptides
 CC generated by these generic sequences are shown in R65150-54. TSARs are
 CC concatenated heterofunctional proteins or peptides, comprising at least
 CC two functional regions - a binding domain with affinity for a ligand and
 CC a second effector peptide portion that is chemically or biologically
 CC active. They may further comprise a linker peptide between the 2 domains.
 CC The oligonucleotides are also designed so that the expressed peptide
 CC contains 2 or 4 cysteine residues positioned in, or flanking, the
 CC conformational rigidity to the peptides. These residues confer some degree of
 CC a TSAR binding domain can be used in vivo to deliver a chemically or
 CC biologically active moiety, eg. metal ion, radioisotope, peptide, toxin
 CC or enzyme, to the specific target or on the cell. They can also replace
 CC the function of macromolecules, eg. monoclonal or polyclonal antibodies
 CC and therefore circumvent the need for complex methods of hybridoma
 CC formation or in vivo antibody production. The TSARs are easily
 CC characterised and have designed activity allowing direct and rapid
 CC detection in a screening process.
 SQ Sequence 114 BP; 0 A; 4 C; 4 G; 4 T;

Query Match 2.2%; Score 35; DB 12; Length 114;
 Best Local Similarity 7.1%; Pred. No. 6.86e-05;
 Matches 8; Conservative 31; Mismatches 74; Indels 0; Gaps 0;
 Db 2 gcnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnn 61
 QY 149 GCCCGGGGCGCATGCTGCGGGTGGGAGCGCGCGCGCTCAGCCCGCCCTCGGCTGC 208
 Db 62 nbnnbntgcnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbntgc 114
 QY 209 TCTCTCTCTCCGCTGGGAGGGCGGTATCTCGGGGCGCTCGCCAGCCCGGC 261

RESULT 11
 ID Q70465 standard; DNA; 114 BP.
 AC Q70465;
 DT 05-APR-1995 (first entry)
 DE Generic DNA sequence to generate a random TSAR peptide library.
 KW TSAR: totally synthetic affinity reagent; synthetic; binding domain;
 KW effector domain; concatenated heterofunctional protein; linker;
 KW direct; rapid; detection; screening; treatment; generic; ss.

OS Synthetic.
 FH Key Location/Qualifiers
 FT misc_feature 55..60
 FT /*tag= a
 FT /note= "this sequence represents 'Z'; Z can be a
 FT sequence of 6, 9 or 12 nucleotides (see
 FT comments)"
 PN WO9418318-A.
 PD 18-AUG-1994.
 PF 01-FEB-1994; U00977.
 PR 01-FEB-1993; US-013416.
 PR 30-DEC-1993; US-176500.
 PR 31-JAN-1994; US-189331.
 PA (UYN-) UNIV NORTH CAROLINA.
 PI Fowlkes DM, Kay BK;
 DR WPI: 94-279739/34.
 DR P-PSDB; R65150 and R65151.
 PT Identifying proteins or peptide(s) which bind a ligand - by
 PT screening a recombinant vector library expressing fusion proteins
 PT comprising a binding domain and an effector domain
 PS Disclosure; Page 35; 255pp; English.
 CC Q70465 is a generic DNA sequence used to generate random TSAR (Totally
 CC represented as follows: X(NNB)6(TGC)(NNB)11Z(NNB)14(TGC)(NNB)3Y. X
 CC and Y are flanking restriction sites (X is not the same as Y) that are
 CC not specified further. Other generic sequences are shown in Q70466-68.
 CC Other specific peptides generated by these generic sequences are shown in
 CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
 CC comprising at least two functional regions - a binding domain with
 CC affinity for a ligand and a second effector peptide portion that is
 CC chemically or biologically active. They may further comprise a linker
 CC peptide between the 2 domains. The oligonucleotides are also designed so
 CC that the expressed peptide contains 2 or 4 cysteine residues positioned
 CC in, or flanking, the unpredicted or variant residues. These residues
 CC confer some degree of conformational rigidity to the peptides. The TSARs
 CC or compps. comprising a TSAR binding domain can be used in vivo to
 CC deliver a chemically or biologically active moiety, eg. metal ion,
 CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
 CC cell. They can also replace the function of macromolecules, eg.
 CC monoclonal or polyclonal antibodies and therefore circumvent the need
 CC for complex methods of hybridoma formation or in vivo antibody
 CC production. The TSARs are easily characterised and have designed
 CC activity allowing direct and rapid detection in a screening process.
 SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 2.2%; Score 34; DB 12; Length 114;
 Best Local Similarity 3.6%; Pred. No. 2.04e-04;
 Matches 4; Conservative 33; Mismatches 75; Indels 0; Gaps 0;
 Db 3 bnnbnnbnnbnnbntgcnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62
 QY 106 GCCTCGGGCACTAACCGAGTGTTCGGGGGGCTGTGAGGGAGGGCCCGGCCCATTCG 165
 Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbntgcnbnnbnnb 114
 QY 166 TGGCGGTGGAGCGCGCGCGGTCTCAGCCCGCGCTCGCTCTCTCTCTCT 217

RESULT 12
 ID Q70467 standard; DNA; 114 BP.
 AC Q70467;
 DT 05-APR-1995 (first entry)
 DE Generic DNA sequence to generate a random TSAR peptide library.
 KW TSAR: totally synthetic affinity reagent; synthetic; binding domain;
 KW effector domain; concatenated heterofunctional protein; linker;
 KW direct; rapid; detection; screening; treatment; generic; ss.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT misc_feature 55..60
 FT /*tag= a
 FT /note= "this sequence represents 'Z'; Z can be a
 FT sequence of 6, 9 or 12 nucleotides (see
 FT comments)"

```
PN W09418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994. U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI; 94-279739/34.
DR P-PSDB; R65153.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 35; 255pp; English.
CC Q70467 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)16(TGC)(NNB)12(NNB)14(TGC)(NNB)1Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in Q70466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active. They may further comprise a linker
CC peptide between the 2 domains. The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or compns. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need for
CC complex methods of hybridoma formation or in vivo antibody production.
CC The TSARs are easily characterised and have designed activity allowing
CC direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 2.2%; Score 34; DB 12; Length 114;
Best Local Similarity 1.9%; Pred. No. 2.04e-04;
Matches 2; Conservative 33; Mismatches 71; Indels 0; Gaps 0;

Db 3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnn 62
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 GAGTGTTCGGGGGCTGTGAGGGAGGGCCCGCGGCCATTGCTCTCTCTCCGCGTGGG 180
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 108
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 181 GCGCGGTCTCAGCCGCGCTCGGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 226
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
ID Q70465 standard; DNA; 114 BP.
AC Q70465.
DT 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc.feature 55..60
FT /*tag= a
FT /*note= "this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)
PN W09418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994. U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
```

```
PI Fowlkes DM, Kay BK;
DR WPI; 94-279739/34.
DR P-PSDB; R65150 and R65151.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 35; 255pp; English.
CC Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)16(TGC)(NNB)11Z(TGC)(NNB)3Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in Q70466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active. They may further comprise a linker
CC peptide between the 2 domains. The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or compns. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need
CC for complex methods of hybridoma formation or in vivo antibody
CC production. The TSARs are easily characterised and have designed
CC activity allowing direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 2.2%; Score 35; DB 12; Length 114;
Best Local Similarity 3.7%; Pred. No. 6.86e-05;
Matches 4; Conservative 33; Mismatches 72; Indels 0; Gaps 0;

Db 3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnn 62
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 114 TGGCGAGCGCTGTGTGCGCGCGCGAGCTCTGTAGCTCAGGACCTCTCTCTCC 55
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 111
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 54 CACGGCGTCCCGGCTGTGTACCCCGTCCGGATCCCGGCGCTCTCTCC 6
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
ID Q70470 standard; DNA; 114 BP.
AC Q70470.
DT 10-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc.feature 55..60
FT /*tag= a
FT /*note= "encoded by Z (see comments)"
PN W09418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994. U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI; 94-279739/34.
DR P-PSDB; R58378.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 36; 255pp; English.
CC Q70470 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
```

CC represented as follows: X(NNB)4(CAC)(NNB)4(CAC)(NNB)8Z(CAC)(NNB)6(CAC)(NNB)8
 CC -(CAC)2(NNB)Y. X and Y are flanking restriction sites (X is not the same
 CC as Y) that are not specified further. The peptides generated by this and
 CC other generic sequences (Q70471-73) have invariant histidine residues
 CC incorporated into variant sequences. TSARS are concatenated
 CC heterofunctional proteins or peptides, comprising at least two functional
 CC regions - a binding domain with affinity for a ligand and a second
 CC effector peptide portion that is chemically or biologically active. They
 CC may further comprise a linker peptide between the 2 domains. The TSARS
 CC or compsns comprising a TSAR binding domain can be used in vivo to
 CC deliver a chemically or biologically active moiety, eg. metal ion,
 CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
 CC cell. They can also replace the function of macromolecules, eg.
 CC monoclonal or polyclonal antibodies and therefore circumvent the need
 CC for complex methods of hybridoma formation or in vivo antibody
 CC production. The TSARS are easily characterised and have designed
 CC activity allowing direct and rapid detection in a screening process.
 SQ Sequence 114 BP; 5 A; 10 C; 0 G; 0 T;

Query Match 2.2%; Score 34; DB 12; Length 114;
 Best Local Similarity 9.8%; Pred. No. 2.04e-04;
 Matches 11; Conservative 29; Mismatches 72; Indels 0; Gaps 0;

Db 3 bnnbnnn 62
 Cp 150 GGCCTCCCTCACAGCCCGCGGACACACTCGGTACTGCGAGGCGCTGTGTGCCAGCC 91
 Db 63 bnnbn 114
 Cp 90 GCGGCAGCTGTGTAGCTCAGGACCTCTCTCAGCCGCGCAGCGCTGTGTGCCCGGT 39

RESULT 15
 ID Q70468 standard; DNA; 114 BP.
 AC Q70468;
 DE 05-APR-1995 (first entry)
 DT Generic DNA sequence to generate a random TSAR peptide library.
 KW TSAR: totally synthetic affinity reagent; synthetic; binding domain;
 KW effector domain; concatenated heterofunctional protein; linker;
 KW direct; rapid; detection; screening; treatment; generic; ss.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT misc_feature 55..60
 FT /*tag= a
 FT /note= "this sequence represents 'Z'; Z can be a
 FT sequence of 6, 9 or 12 nucleotides (see
 FT comments)"
 PN W09418318-A.
 PD 18-AUG-1994.
 PF 01-FEB-1994; U00977.
 PR 01-FEB-1993; US-013416.
 PR 30-DEC-1993; US-176500.
 PR 31-JAN-1994; US-189331.
 PA (UYN-) UNIV NORTH CAROLINA.
 PI Fowlkes DM, Kay BK;
 DR WPI: 94-279739/34.
 DR P-PSDB: R65154.
 PT Identifying proteins or peptide(s) which bind a ligand - by
 PT screening a recombinant vector library expressing fusion proteins
 PT comprising a binding domain and an effector domain
 PS Disclosure: Page 35; 253pp; English.
 CC Q70468 is a generic DNA sequence used to generate random TSAR (Totally
 CC Synthetic Affinity Reagents) peptides. This generic formula can also be
 CC represented as follows: X(NNB)11(TGC)(NNB)8Z(NNB)7(TGC)(NNB)10Y. X
 CC and Y are flanking restriction sites (X is not the same as Y) that are
 CC not specified further. Other generic sequences are shown in Q70466-68.
 CC Other specific peptides generated by these generic sequences are shown in
 CC R65151-54. TSARS are concatenated heterofunctional proteins or peptides,
 CC comprising at least two functional regions - a binding domain with
 CC affinity for a ligand and a second effector peptide portion that is
 CC chemically or biologically active. They may further comprise a linker
 CC peptide between the 2 domains. The oligonucleotides are also designed so
 CC that the expressed peptide contains 2 or 4 cysteine residues positioned

CC in, or flanking, the unpredicted or variant residues. These residues
 CC confer some degree of conformational rigidity to the peptides. The TSARS
 CC or compsns comprising a TSAR binding domain can be used in vivo to
 CC deliver a chemically or biologically active moiety, eg. metal ion,
 CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
 CC cell. They can also replace the function of macromolecules, eg.
 CC monoclonal or polyclonal antibodies and therefore circumvent the need
 CC for complex methods of hybridoma formation or in vivo antibody
 CC production. The TSARS are easily characterised and have designed activity
 CC allowing direct and rapid detection in a screening process.
 SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 2.2%; Score 34; DB 12; Length 114;
 Best Local Similarity 4.5%; Pred. No. 2.04e-04;
 Matches 5; Conservative 32; Mismatches 75; Indels 0; Gaps 0;

Db 3 bnnbnnn 62
 Cp 177 CTCCACCGCCAGCAATGGCGCGCGGCGCCCTCCCTCACAGCCCGCGGAACACTCGGT 118
 Db 63 bnnbn 114
 Cp 117 TAGTCCGAGCGCTGTGTGTGCCAGCGCGCGAGCTGTGTAGCTCAGGACC 66

Search completed: Thu Nov 6 11:10:04 1997
 Job time : 197 secs.

[W][A][T][E][R][M][A][N]

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by Intelligenetics, Inc.

MPSrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Nov 6 11:10:33 1997; MasPar time 829.51 Seconds
Tabular output not generated. 969.077 Million cell updates/sec

Title: >US-08-842-827-3
Description: (1-1566) from US08842827.seq
Perfect Score: 1566
N.A. Sequence: 1 CCTGTGGAGAGAGCGCGG.....CCAAAAA.....1566
Comp: GGACACCTCTCTCGCGGCC.....GGTTTTTTTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 707517 seqs, 256659390 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: EST-STS
9:EST9 10:EST10 11:EST11 12:EST12 13:EST13 14:EST14
15:EST15 16:EST16 17:EST17 18:EST18 19:EST19 20:EST20
21:EST21 22:EST22 23:EST23 24:EST24 25:EST25 26:EST26
27:EST27 28:EST28 29:EST29 30:EST30 31:EST31 32:EST32
33:EST33 34:EST34 35:EST35 36:EST36 37:EST37 38:EST38
39:EST39 40:EST40 41:EST41 42:EST42 43:EST43 44:EST44
45:EST45 46:EST46 47:EST47 48:EST48 49:EST49 50:EST50
51:EST51 52:EST52 53:EST53 54:EST54 55:EST55 56:EST56
57:EST57 58:EST58 59:EST59 60:EST60 61:EST61 62:EST62
63:EST63 64:EST64 65:EST65 66:EST66 67:EST67 68:EST68
69:EST69 70:EST70 71:EST71 72:EST72 73:EST73 74:EST74
75:EST75 76:EST76 77:EST77 78:EST78 79:EST79 80:EST80
81:EST81 82:EST82 83:EST83 84:EST84 85:EST85 86:EST86
87:EST87 88:EST88 89:EST89 90:EST90 91:EST91 92:EST92
93:EST93 94:EST94 95:EST95 96:EST96 97:EST97 98:EST98
99:EST99

EST-STS-TWO
100:EST100 101:EST101 102:EST102 103:EST103 104:EST104
105:EST105 106:EST106 107:EST107 108:EST108 109:EST109
110:EST110 111:EST111 112:EST112 113:EST113 114:EST114
115:EST115 116:EST116 117:EST117 118:EST118 119:EST119
120:EST120 121:EST121 122:EST122 123:EST123 124:EST124
125:EST125 126:EST126 127:EST127 128:EST128 129:EST129
130:EST130 131:EST131 132:EST132 133:EST133 134:EST134
135:EST135 136:EST136 137:EST137 138:EST138 139:EST139
140:EST140 141:EST141 142:EST142 143:EST143 144:EST144
145:EST145 146:EST146 147:EST147 148:EST148 149:EST149
150:EST150 151:EST151 152:EST152 153:EST153 154:EST154
155:EST155 156:EST156 157:EST157 158:EST158 159:EST159
160:EST160 161:EST161 162:EST162 163:EST163 164:EST164
165:EST165 166:EST166 167:EST167 168:EST168 169:EST169
170:EST170 171:EST171 172:EST172 173:EST173 174:EST174

175:EST175 176:EST176 177:EST177 178:EST178 179:EST179
180:EST180 181:EST181 182:EST182 183:EST183 184:EST184
185:EST185 186:EST186 187:EST187 188:EST188 189:EST189
190:EST190 191:EST191 192:EST192 193:EST193 194:EST194
195:EST195 196:EST196 197:EST197 198:EST198

Statistics: Mean 11.818; Variance 2.886; scale 4.094

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description	Pred. No.
C 1	528	33.7	587	75	H97570	yx03b06.s1 Homo sapie	0.00e+00
C 2	515	32.9	545	137	AA043085	zk48b02.r1 Soares pre	0.00e+00
C 3	507	32.4	552	159	AA036943	zk30f04.s1 Soares pre	0.00e+00
C 4	501	32.0	546	171	W67667	zk38a11.s1 Soares pre	0.00e+00
C 5	461	29.4	465	160	AA040858	zk48b02.s1 Soares pre	0.00e+00
C 6	413	26.4	472	114	W04968	zk43g09.r1 Soares pre	0.00e+00
C 7	414	26.4	490	92	N31047	yx51h07.r1 Homo sapie	0.00e+00
C 8	406	25.9	446	171	W67666	zk38a11.r1 Soares pre	0.00e+00
C 9	405	25.9	456	169	W45106	zk21a10.r1 Soares pre	0.00e+00
C 10	406	25.9	474	113	W01375	zk20f11.r1 Soares pre	0.00e+00
C 11	395	25.2	422	169	W45155	zk21a10.s1 Soares pre	0.00e+00
C 12	381	24.3	456	67	H68363	yr02f11.r1 Homo sapie	0.00e+00
C 13	380	24.3	471	189	AA152123	zk14g02.r1 Soares pre	0.00e+00
C 14	377	24.1	479	74	H90961	yu86e01.r1 Homo sapie	0.00e+00
C 15	376	24.0	415	151	W24605	zk63f08.r1 Soares pre	0.00e+00
C 16	374	23.9	423	78	R97295	yq74h06.r1 Homo sapie	0.00e+00
C 17	357	22.8	452	85	H54373	yq93g09.r1 Homo sapie	0.00e+00
C 18	357	22.8	471	85	H57543	yr05d10.r1 Homo sapie	0.00e+00
C 19	357	22.8	480	169	W45113	zz21c09.r1 Soares pre	0.00e+00
C 20	354	22.6	404	62	H17855	ym36b12.s1 Homo sapie	0.00e+00
C 21	347	22.2	438	4	T70188	yc18c09.s1 Homo sapie	0.00e+00
C 22	347	22.2	473	85	H54296	yq93g09.s1 Homo sapie	0.00e+00
C 23	343	21.9	445	85	H57544	yr05d10.s1 Homo sapie	0.00e+00
C 24	342	21.8	356	165	C17623	Human placenta CDNA 5	0.00e+00
C 25	334	21.3	386	5	T70311	yd6h07.s1 Homo sapie	0.00e+00
C 26	322	20.6	331	158	AA033777	zk19g11.r1 Soares pre	0.00e+00
C 27	321	20.5	351	85	H57213	yr08d07.r1 Homo sapie	0.00e+00
C 28	319	20.4	332	158	AA033778	zk19g11.s1 Soares pre	0.00e+00
C 29	314	20.1	357	159	AA037575	zk34f08.r1 Soares pre	0.00e+00
C 30	315	20.1	408	7	H80888	yd6h07.r1 Homo sapie	0.00e+00
C 31	310	19.8	357	64	H26307	yl51f02.s1 Homo sapie	0.00e+00
C 32	308	19.7	386	159	AA036941	zk30f04.r1 Soares pre	0.00e+00
C 33	304	19.4	413	67	H68364	yr02f11.s1 Homo sapie	0.00e+00
C 34	298	19.0	422	4	T69858	yc18c09.r1 Homo sapie	0.00e+00
C 35	287	18.3	304	198	H96367	yw61f11.s1 Soares pre	0.00e+00
C 36	270	17.2	279	108	HSC37E112	H. sapiens partial cd	0.00e+00
C 37	269	17.2	315	9	T90107	yd9g02.s1 Homo sapie	0.00e+00
C 38	265	16.9	287	142	N86994	L1773F Fetal heart, L	0.00e+00
C 39	257	16.4	351	36	R07363	ye96g08.s1 Homo sapie	0.00e+00
C 40	249	15.9	252	189	AA150023	zk14g01.s1 Soares pre	0.00e+00
C 41	247	15.8	325	128	W92816	zk48b08.s1 Soares pre	0.00e+00
C 42	248	15.8	331	37	R09946	yf30f03.r1 Homo sapie	0.00e+00
C 43	237	15.1	247	108	HSC36C022	H. sapiens partial cd	0.00e+00
C 44	234	14.9	511	154	AA008988	mg99e04.r1 Soares mou	0.00e+00
C 45	226	14.4	354	73	H90307	yu86e01.s1 Homo sapie	0.00e+00

ALIGNMENTS

RESULT 1
LOCUS H97570 587 bp mRNA EST 12-DEC-1995
DEFINITION yx03b06.s1 Homo sapiens CDNA clone 260627 3'
ACCESSION H97570
NID G1118455
KEYWORDS EST.
SOURCE human clone-260627 primer=ml3 -40 forward library=Soares melanocyte

2N8HM vector-prrT73D (Pharmacia) with a modified polylinker host-DH10B (ampicillin resistant) Isceti-I Not I RsaI2-Eco RI Male. The last strand cDNA was primed with a Not I - oligo(dT) primer (5'-TGTTACCAATCGAATGGAGCGCCGAGTGTGTTTTTTT-3'), double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified prT73 vector (Pharmacia). Library constructed by Bento Soares and M. Fatima Bonaldo. RNA from foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P. Albino.

ORGANISM

Homo sapiens
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Hominoidei; Homo.
1 (bases 1 to 587)
Hillier, L., Clark, N., Dubucque, T., Ellison, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 351
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES

```

FEATURES
  source
    1..587
      Location/Qualifiers
        organism="Homo sapiens"
        /clone="260627"
        mRNA
        <1..>587
      BASE COUNT
        173 a 117 c 120 g 174 t 3 others
      ORIGIN
        Query Match 33.7%; Score 528; DB 75; Length 587;
        Best Local Similarity 96.6%; Pred. No. 0.00e+00;
        Matches 565; Conservative 0; Mismatches 16; Indels 4; Gaps 4;

```

Db	1	gggtcccccacatctttaatatgatattatattataatttaataccaacatcatctatcatagaagca	60
Cp	1548	GGTTCGCCATACATTTTAAATGTATTATATTTAAATCAAACATCATTCATACAAAGCA	1489
Db	61	tattacatacatgtttcatacataaagcattacatcttttttaataaaaaatgtatcacagtgg	120
Cp	1488	TATTACATACATGTTTATACATAAGCATTACATTTTTTTTAAATAAAATGTATACAGTGG	1429
Db	121	ggcactgcttttggggaagccttgaggtttttttaaataagtttagagagctatagataaac	180
Cp	1428	GGCACGTGTTTGGTGTGAAGCCTGGAGTTTTTTTTTAATGAGTTTAGAGCTATTAGATAACC	1369
Db	181	actgaagttaaaggttaactatgtacacacaaaagtg-gcatccaaagaggcataagcagcagca	239
Cp	1368	ACTGAGTTTAAAGTAATATGTATACACACAAAGTGTGCATCCAAAGGCGATATCAGCAGCA	1309
Db	240	gaagtcctttaaaggcttgcacaccagggaagaagatgcactctcttgcttgtggaatc	299
Cp	1308	GAAGTCCTTTAAAGCTGTGTACACAGGAAGAAGATGCATCCTCTTGCTTGTGGCAATC	1249
Db	300	atttctctttgaaaaacagccagcttccacctgggcacccctgcctgccttcaagggctggg	359
Cp	1248	ATTTTCTTTAGAAAAACAGCCAGCTTCACCTGGGCAACCTGCTGCCCTTTCAGGCTGG-	1190
Db	360	tgattgctcggatagtgattccccagttgttggttttcatgcagagttgtatgagagtcc	419

Cp	1189	TGATTGCTCGGATAGTGAATCCAGTTGGTGTTTCATGCAGAGTTGTATGAGAGTCC	1130
Db	420	tctctccttccctctttaagaagatctttctctgagaataacgcatacacatacacagca	479
Cp	1129	TCCTCTTTTCTTCTTTAAAGAAGTCTTCTTTTGAGAANAATCCGATACATATACAGCA	1070
Db	480	actaatattgcaaccaggagctcccgaatgatgcttcagtgcaacaacatcgctccccagtggg	539
Cp	1069	ACTAATATTGCAACCAGAGCTCCCTGAATGAGTCGAGTCAACACATCGCTCC-AGTGGTG	1011
Db	540	gtttataatccgnactccngaaaggnccccaataaatggggaacgg	584
Cp	1010	-TTTTTAATCAGAAAATCCGAGAAAGGCCACATAAATGGATACGG	957

RESULT 2

RESULTS	LOCUS	DEFINITION	545 bp	mRNA	EST	04-SEP-1996
		AA043085	Soares pregnant uterus	NBHPU	Homo sapiens cDNA clone	
		486027 5'				
	ACCESSION	AA043085				
	NID	91520979				
	KEYWORDS	EST.				
	SOURCE	human.				

ORGANISM	Homo sapiens
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

Eukaryotes: mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Euthera; Primates; Catarrhini; Hominoidea; Homo
 1 (bases 1 to 545)
 REFERENCE
 AUTHORS
 Hillier, L., Clark, N., Dubucq, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, C., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE The WashU-Merck EST protect

FILE
JOURNAL
COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 485.

FEATURES

```

source
location/Qualifiers
1..545
/organism="Homo sapiens"
/note="Organ: uterus; Vector: pT73-Pac; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AAGTGAAGAAATTCGCCGCCCTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

```

mrna	BASE COUNT	ORIGIN
159 a	116 c	105 g
163 t	105 g	163 t
2 others	105 g	163 t

```

Query Match      32.9%; Score 515; DB 137; Length 545;
Best Local Similarity 98.7%; Pred. No. 0.00e+00;
Matches 540; Conservative 0; Mismatches 3; Indels 4; Gaps 4;

Db 1 cccacactgcgaatttggctgttgcgtatccattatgtgggcctttctcgnngttct 60

```


REFERENCE
AUTHORS

Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 546)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultnan,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.

TITLE
The WashU-Merck EST Project

JOURNAL
Unpublished (1995)

COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 634 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 321.

FEATURES
source

Location/Qualifiers
1..546

```

/organism="Homo sapiens"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGGCGCARTTTTTTTTTTTTTTTT 3'] ,
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbHL19W."
/clone_lib="Soares fetal heart NbHH19W"
/clone_lib="342908"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
complement(<1..>546)
BASE COUNT      159 a   105 c   121 g   159 t    2 others
ORIGIN

Query Match          32.0%; Score 501; DB 171; Length 546;
Best Local Similarity 97.6%; Pred. No. 0.00e+00;
Matches 523; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

Db      2      ttatcataagcattacattttttaataaaaatgtatcacaggctgggacctgttgg 61
Cp      1475   TTTATACATAAGCATTCATTTTTAAATAAAAATGTATACAGGTGGGCACGTGTTGG 1416

Db      62      tggaaggcttgaggttttttaagttagtgactattagataaccactgagttaaagg 121
Cp      1415   TGGAAGGCTTGGAGTGTTTTAAATGAGTTTAGAGCTATTAGATACCACCTGAGTTAAGG 1356

Db      122      taactatgcacacaaagtggcnatcccaaggcgcatagcagcagagaagtctttaaag 181
Cp      1355   TAACTATGTACACAAAAAGTGTGCATCCAAGAGGCATATCACGACGAGAAGCTCTTAAG 1296

Db      182      gcttgtacacagaagaagatgatcctcttgctgtgg-aatcatttccttaga 240
Cp      1295   GCTTGTCACCCAGGAAGAAGATGATCCTCTGTGCTGTGGCAATCATYTTTCCTTAGA 1236

Db      241      aaacaggccagcttcacctggcaccctcgctgctttcaaggctggtagtgcctcgata 300
Cp      1235   AAACAGGCCAGCTTCACCTGGGCACCCCTGCTGCCCTTCAAGGCTGGTGTGCTCGGATA 1176

Db      301      gtgattcccagttgtgtgttoaatgcagagttgtatgagagtcctoctctttcttc 360
Cp      1175   GTGATTCCCAGTTGTGTGTGTTTCATGAGAGTGTGTATGAGAGTCTCTCTCTTTCTTTC 1116

```

```
Db 61 tattacacatgtttatcacataagcattacatttttttaataaaaaatgtatcacagtgg 120
|||||
Cp 1488 TATTACATACATGTTTATACATAAGCATTACATTTTATTAATAAAATGTATACAGTGG 1429
|||||
Db 121 ggcactgttttgggagagccttgagattttttaaagatttttagagctattagataacc 180
|||||
Cp 1428 GGCACGTGTTTGGTGGAGGCTTGGAGTTTTTTTAAATGAGTTTAGAGCTATTAGATAACC 1369
|||||
Db 181 actgagtaaaaggttaactatgtcacacaaagtgtcatccaagagcagcagcagca 240
|||||
Cp 1368 ACTGAGTTAAAGGTAACTATGTACACACAAGTGTGTCATCCAAGAGCATATACAGACGA 1309
|||||
Db 241 gaagctttttaaagccttgcacacaggaagaagatgcattcctcttgccttgggcaatc 300
|||||
Cp 1308 GAAGCTCTTTAAAGCCTTGTACACAGGAAGAAGATGCATCCTCTTGCCTTGTGCAATC 1249
|||||
Db 301 attttcctttagaacaaacagccagccttcaacctggggcccttgcctttcaaggctgg 360
|||||
Cp 1248 ATTTCCTCTTTAGAAACAGGCCAGCTTCACCTGGGCACCCCTGCTGCCCTTCAAGGCTGGT 1189
|||||
Db 361 gattgctcggatagatgattcccaattgttgggtttcatgcagagttgtatgagagtcct 420
|||||
Cp 1188 GATTGCTCGGATAGTATCCCACTTGTGGTGTTTCATGCGAGAGTGTATGAGAGTCTCT 1129
|||||
Db 421 cctcttttcttctttaaagaagttcttcttcttgaagaataccg 465
|||||
Cp 1128 CCTCTTTCTCTTTAAAGAAGTCTCTTCTTTGAAGAAATCCG 1084
|||||

RESULT 6
LOCUS W04968 472 bp mRNA EST 23-APR-1996
DEFINITION za43q09.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
295360 5' similar to WP:T28D9.3 CE02068 ;.
ACCESSION W04968
NID g1277708
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 472)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlffing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: mob.REGA+ET
High quality sequence stop: 339.
Location/Qualifiers
1..472
/organism="Homo sapiens"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACTGGAAGAATTAATAAGATCTTTTATTTTATTTT 3'] ,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
```

```
/clone="295360"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
<1..>472

mrna 120 a 102 c 98 g 152 t
BASE COUNT
ORIGIN

Query Match 26.4%; Score 413; DB 114; Length 472;
Best Local Similarity 98.6%; Pred. No. 0.00e+00;
Matches 429; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

Db 37 tccgttattattcttgagaaacccctgtctgtttactgttaacctttttgacctcaaatcc 96
|||||
QY 549 TCCTCTATTATTCTGGAGAAACCCCTGCTGTTCGTACGTAAACCTTTTGACACTCAATTC 608
|||||
Db 97 ttatcagggaataacacacacacacacacacacacacacacacacacacacacacacacac 156
|||||
QY 609 TTTATCAGTAATAACTACATAGGCACATATTACAAAGCCATTGGAACTTTTATTATTGGT 668
|||||
Db 157 gcagctcgtactcagccctcagcagcatttgcacagatttcaagatttcaagcagactcgccc 216
|||||
QY 669 GCAGCTCTAGTCTAGTCCCTGACTGCATTT-GCCAAAGTATTCAATAGGCAGACTGCGGCC 727
|||||
Db 217 tcacttttggatgtttgtgctcagattgttcaaaaatcaactgcaagcgatggtttaca 276
|||||
QY 728 TCACCTCTCTGGATGTTGTGTCATCCAGATTGGTCAAAAATCAACTGCA-GGATGCGTTTACA 786
|||||
Db 277 ttgaatactacatatgtcgaggggaatgcagaagagtgtaaggaggcaggttgcctctct 336
|||||
QY 787 TTGAATACTACATATGTCGAGGGAATGCAGAAAGAGTTAAAGGAAGGCAGGTGTGCTCTCT 846
|||||
Db 337 attcaggccactcttcttccatgctactgctgctgttcttggcactttattctcaag 396
|||||
QY 847 ATTCAAGCCACTCTCTCTCTTTCATGTACTGCATGCTGTGTGTGGCACTTTATCTTCAAG 906
|||||
Db 397 ccaggatgaaggagagactggcagagactcttcagccccacacactgcaatttggctcttg 456
|||||
QY 907 CCAGGATGAGGAGGAGACTGGGCAAGACTCTACGCCCCACACTGCAATTGGTCTTGTGTTG 966
|||||
Db 457 ccgctatccatttatg 471
|||||
QY 967 CCGTATCCATTATG 981
|||||

RESULT 7
LOCUS N31047 490 bp mRNA EST 10-JAN-1996
DEFINITION yx51h07.r1 Homo sapiens cDNA clone 265309 5' similar to WP:T28D9.3
CE02068 ;.
ACCESSION N31047
NID g1151446
KEYWORDS EST.
SOURCE human clone=265309 primer=T7 library=Soares melanocyte 2NBHM
vector=pT7T3D (Pharmacia) with a modified polylinker host=DH10B
(ampicillin resistant) RsaI=Not I RsaI2=Eco RI Male. 1st strand
cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGGCCAGCTTTTATTTTATTTTATTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I and Eco
RI sites of a modified pT7T3 vector (Pharmacia). Library
constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal
foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P.
Albino.
ORGANISM Homo sapiens
Eukaryotes; Metazoa; Eumetazoa; Bilateria; Coelomata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 490)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlffing,T., Soares,M., Tan,F.,
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: mob.REGA+ET
High quality sequence stop: 339.
Location/Qualifiers
1..472
/organism="Homo sapiens"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACTGGAAGAATTAATAAGATCTTTTATTTTATTTT 3'] ,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
```

TITLE	COMMENT
<p> Trevaaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R. The WashU-Merck EST Project Unpublished (1995) Contact: Wilson RK WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu High quality sequence stops: 385 Source: IMAGE Consortium, LNL This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. FEATURES source 1..490 Location/Qualifiers /organism="Homo sapiens" /clone="265309" <1..>490 BASE COUNT 131 a 107 c 111 g 138 t 3 others ORIGIN Query Match 26.4%; Score 414; DB 92; Length 490; Best Local Similarity 97.1%; Pred. No. 0.00e+00; Matches 476; Conservative 0; Mismatches 4; Indels 10; Gaps 10 Db 1 gncgcctactcttgatgtgttcgcgataccagatggtgcacaaatcaactgcagcgatgg 60 Qy 722 GCGCCCTCACATCTTGGATGTTGTGTATCCAGATGTGGTCAAAAATCAACTGCAGCGATGG 781 Db 61 ttacatgaactacacatatctcagggaaatgcagaaanagttaagaaagcaggttgct 120 Qy 782 TTACATTTGAATACTACATATGTCGAGGGAAATGCAGAAAGAGTTAAGGAAGCGAGTTGTC 841 Db 121 cttctattcaggccactcttcgtttccatgtaactgcgtctgtttgtggcactttatct 180 Qy 842 CTTCTATTACGGCCACTCTTCGTTTTCATGTACTGCATGCTGTTTGTGGCACTTTATCT 901 Db 181 tcaagcagagatgaaggagactggcgaagactcttcagcccaacacatgcaattggctct 240 Qy 902 TCAAGCCAGGATGAAGGGAGACTGGGCAAGACTCTTACGCCCCACACTGCAATTTGGTCT 961 Db 241 tgttcgcgtatccatttatgtgggcctttctcgagttcttgattataaacaccactggag 300 Qy 962 TGTTCGGGTATCCATTATATGTGGGCCCTTTCTCGAGTTTCTGATATATAACACCACCTGGAG 1021 Db 301 cgatgttgactggactcatttcaggagctctcgttgcgaattagttgctgatatgt 360 Qy 1022 CGATGTGTGTGACTGGACTCATTCAGGGAGGCTCTGTTTGCATATATTAGTTCGTGTATATGT 1081 Db 361 atcggattctctcaaaagaagaactctcttttaagaaagaaagggaggactctcca 420 Qy 1082 ATCGGATTTCTCAAGAAAGAAAGACTTC-TTTTAAAGAAAGAAAG-AGGAGGACTCTC-A 1138 Db 421 tacaactctggcattgaaacaccacacaaactgggaatncactaatccggagcaatccacc 480 Qy 1139 TACAACCTC-TGC-ATG-AAACACCACAACTGGGAAT-CACATA-TCCG-AGCAATC-ACC 1191 Db 481 agccttgaan 490 Qy 1192 AGCCTTGAAA 1201 RESULT 8 LOCUS W67666 446 bp mRNA EST 15-OCT-1996 DEFINITION Z438a11.r1 Soares fetal heart NDH19W Homo sapiens cDNA clone 342908 5'. ACCESSION W67666 NID g1376761 KEYWORDS EST. SOURCE human.</p>	

```

|||||
QY 1195 CTTGAAAGGACAGAGGTGCCAGGTGAAGTGGCTGTTTCTTAAAGGAAAATGATTGC 1254
|||||
Db 361 cacaagcaagaagatgatcttcttctctgtgtacaaagcctttaagactctggc 420
|||||
QY 1255 CACAAGGCAA-GA-GGATGCACTTCTTCTCTGGTGACAAAGCCTTTAAAGACTTCG-C 1311
|||||
Db 421 tgctgtatgcctctt 436
|||||
QY 1312 TGCTGATATGCTCTT 1327
|||||

RESULT 9
LOCUS w45106 456 bp mRNA EST 10-OCT-1996
DEFINITION zc21a10.r1 Soares senescent fibroblasts NBHSF Homo sapiens cDNA
clone 322938 5'.
ACCESSION W45106
NID g1329189
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 456)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 728 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 435.
FEATURES
source
1..456
/organism="Homo sapiens"
/note="Vector: p7T3D (Pharmacia) with a modified
polylinker V.TYPE: phagemid; Site_1: Not I; Site_2: Eco
RI; TGTTACCAATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT
3'), double-stranded cDNA was size selected, ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified p7T3
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo."
/clone="322938"
/clone_lib="Soares senescent fibroblasts NBHSF"
/tissue_type="senescent fibroblast"
/lab_host="DH10B (ampicillin resistant)"
<1..>456

BASE COUNT 124 a 98 c 107 g 126 t 1 others
ORIGIN

Query Match 25.9%; Score 405; DB 169; Length 456;
Best Local Similarity 98.6%; Pred. No. 0.00e+00;
Matches 436; Conservative 0; Mismatches 1; Indels 5; Gaps 4;

Db 15 cttatcttcagccagatgaaggagactgggcaaacctcttacgcccacactgcaa 74
|||||
QY 894 CTTTATCTCAAGCCAGGATGAAGGAGACTGGGCAAGACTCTTACGCCCCACACTGAA 953
|||||
Db 75 ttgtgtctgtgcgtatccattatgtgggctttctcgtatctgtattataaacac 134
|||||
QY 954 TTRGGTCTTGTTCGGTATCCATTATATGTGGGCTTCTCGAGTTTCTGATTATAAACAC 1013
|||||
Db 135 cactggagcatgtgtgactgactcattcaggagctctgttgcaatattagttgct 194
|||||
QY 1014 CACTGGAGCATGTGTGACTGGACTCATTTCAGGGAGCTCTGGTTGCAATATTAGTTGCT 1073
|||||
Db 195 gtatatgtatcggtattcttcaagaagaactctctttaagaagaagaagaggagac 254
|||||
QY 1074 GTATATGTATCGGATTCTTCAAGAAGAAGAACTCTTTTAAAGAAAGAGGAGGAC 1133
|||||
Db 255 tctcatacaactctgcatgaacacacacacacacacacacacacacacacacacacac 314
|||||
QY 1134 TCTCATACAACCTCTGCATGAACACCAACAACCTGGGAATCATTATCCGACCAATCACCAG 1193
|||||
Db 315 ccttgaaaaagcagcagggtgccagggtgaagctggcctgtttcttaaggaaaaatgat 374
|||||
QY 1194 CCTTGAAG-GGCACAGGGTGCCAGGTGAAGCTGGCCT-GTTTTCTTAAAGGAAAATGAT 1251
|||||
Db 375 tgcacaaggaagagatgcattcttctcctgggtgtgtacaaagcctttaagacttc 434
|||||
QY 1252 TGCACAAAGCAAGAGATGCATCTTCTTCCCTGG--TGACAAAGCCTTT-AAAGACTTC 1308
|||||
Db 435 tgctgtctatgcctctcttggga 456
|||||
QY 1309 TGCTGCTGATATGCTCTTGGGA 1330
|||||

RESULT 10
LOCUS W01275 474 bp mRNA EST 18-APR-1996
DEFINITION za40f11.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
295053 5' similar to contains element MER12 repetitive element ;.
ACCESSION W01275
NID g1273255
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 474)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: EtBprimer
High quality sequence stop: 429.
FEATURES
source
1..474
/organism="Homo sapiens"
/note="Organ: Liver and Spleen; Vector: p7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACGGAAGAATTAATAAGATCTTTTCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
/clone="295053"
/clone_lib="Soares fetal liver spleen INFLS"

```

/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
<1..>474
BASE COUNT 133 a 96 c 101 g 138 t 6 others
ORIGIN

Query Match 25.9%; Score 406; DB 113; Length 474;
Best Local Similarity 96.7%; Pred. No. 0.00e+00;
Matches 466; Conservative 0; Mismatches 6; Indels 10; Gaps 10;

Db 1 gagactgggcaagactcttacgccccacacgcaatttggtctgttgccgtatccattt 60
QY 919 GAGACTGGGCAAGACTCTTACGCCCCACACACTGCAATTTGGTCTGTTCGGTATCCATTT 978
Db 61 atgtggccttctcaggtcttgattatataaacacacacgagcgatgtgtgactggac 120
QY 979 ATGTGGCCCTTCTCAGATTTCTGATTATAAACACCACCTGGACGATGTGTGACTGGAC 1038
Db 121 tcattcaggagctctggttgcaattagttgctgtatgtatcgatttcttcaag 180
QY 1039 TCATTGAGGAGCTCTGGTGGCAATATTAGTTCTGTATATGTATCGGATTTCTTCAAG 1098
Db 181 aaagaactctttaaagaagaagaagagagagactctctacacactctgcataaacac 240
QY 1099 AAAGAAGCTCTTTTAAAGAAAGAAAGAGAGGAGACTCTCATACAACCTCTGCGATGAACAC 1158
Db 241 caacaactgggaatacctatccagagcaataccacagccttgaaaggcagggcgccag 300
QY 1159 CAACAAGTGGGAATCACTATCCGAGCAATACCAGGCTTGAAGAGGAGGAGGCGCCAG 1218
Db 301 gtgaagctggcctgtttctcaaggaaagattggtgcccaagn-aagaggatgcactctt 359
QY 1219 GTGAAGCTGGCCTGTTTCTTAAGGAAATGATGGCCACAGGACAGGATGCATCTTT 1278
Db 360 ctctcctggtgn-aagcctttaaagactn-tgctgtgt-tatnc-tcttgatgcacact 415
QY 1279 CTTCCTGGTGTACAAGCCTTTAAAGACTTCTGCTGTGATATGCCCTTCTGGATGCACACT 1338
Db 416 ttgtgtacatagtac-tttaa-tcagtggtta-ctaatan-tctaaa-tcattaaaa 470
QY 1339 TTGTGTACATAGTTACCTTTAACTCAGTGGTATCTATATAGCTCTAAACTCATTAATA 1398
Db 471 aa 472
QY 1399 AA 1400

RESULT 11 W45155 422 bp mRNA EST 10-OCT-1996
LOCUS zc21a10.s1 Soares senescent fibroblasts NbHSF Homo sapiens cDNA
DEFINITION clone 322938 3'.
ACCESSION W45155
NID g1329246
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 422)
REFERENCE Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
AUTHORS Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.

TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800

Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 728 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 160.
Location/Qualifiers
1..422
/organism="Homo sapiens"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco
RI; TGTACCAATCTGAAGTGGAGGCGGCATTTTTTTTTTTTTT
3'), double-stranded cDNA was size selected, ligated to
Eco RI adapters (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pT73
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo."

FEATURES
source

1..422

/clone_lib="Soares senescent fibroblasts NbHSF"
/tisue_type="senescent fibroblast"
/lab_host="DH10B (ampicillin resistant)"
complement(<1..>422)

BASE COUNT 124 a 73 c 90 g 133 t 2 others
ORIGIN

Query Match 25.2%; Score 395; DB 169; Length 422;
Best Local Similarity 98.3%; Pred. No. 0.00e+00;
Matches 415; Conservative 0; Mismatches 4; Indels 3; Gaps 3;

Db 1 ggtctccatcacatttaataatgatcatatttaatacaaacatcatcattcagaagca 60
Cp 1548 GGTTCTCCCATACATTTAATATGATTATATTAAATCAAAACATCATTCATAGAAAGCA 1489
Db 61 tattacatacatgtttatacataaagcattacatttttttaataaaaaatgtatcacaggtgg 120
Cp 1488 TATTACATACATGTTTATACATAAGCATTTACATTTTAAATAAANAATGTATACAGGTGG 1429
Db 121 ggcactgttttggtggaagcgttgaggttttttaagttagctattagataaac 180
Cp 1428 GGCACGTGTTTGGTGGAGGCTTGGAGTTTTTTTAAATGAGTTAGAGCTATTAGATAACC 1369
Db 181 actgagttaaagtaactatgtacacacaaagtgatcccaagagcagcagcagca 240
Cp 1368 ACTGAGTTAAAGGTAACACTATGTACACACAAAGTGTGCATCCAAAGAGGCATATCAGCAGCA 1309
Db 241 gaagtcctttaaaggctgtgtacaccagggaagaagatgcatcctcttgccttgggaatc 300
Cp 1308 GAAGTCTTTAAAGGCTTGTACACCAGGAAGAAAGATGCATCCTCTTGCCTTGTGGCAATC 1249
Db 301 atttccnttagaaaaaacaggccagcttccactggcaccctgctgccttttcaaggctgg 360
Cp 1248 ATTTTCCTTTAGAAAAACAGGCCAGCTTCACTCCCTGGGACCCCTGCTGCTTTT-CAAGGCTGG 1190
Db 361 tggattgtcgtatagtgattcccgatccaggtgttggttttcacagcagagttgtatagact 420
Cp 1189 TG-ATTGCTCGGATAGTG-ATTCCAGATTGTGGTGTTCATCGACAGATTGTATGAGACT 1132
Db 421 cc 422
Cp 1131 CC 1130

RESULT 12 H68363 456 bp mRNA EST 18-OCT-1995
LOCUS YR82f11.r1 Homo sapiens cDNA clone 211821 5'.
DEFINITION H68363
ACCESSION H68363
NID g1027103
KEYWORDS EST.
human clone=211821 primer=M13Rp1 library=Soares fetal liver spleen
INFLS vector=pT73D (Pharmacia) with a modified polylinker
host=DH10B (ampicillin resistant) Rsite1=Pac I Rsite2=Eco RI Liver

(Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal heart library, Soares fetal heart NBHL19W.
/clone="308295"
/clone_lib="Soares fetal lung NBHL19W"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
<1..>415

BASE COUNT 124 a 87 c 86 g 116 t 2 others
ORIGIN

Query Match 24.0%; Score 376; DB 151; Length 415;
Best Local Similarity 98.1%; Pred. No. 0.00e+00;
Matches 408; Conservative 0; Mismatches 3; Indels 5; Gaps 4;
Db 1 gaggttctgattataaacacacactggagcgatgtgttgactggaactcattcaggagagctc 60
QY 994 GAGTTTCTGATTATAAACACACTGGAGCGATGTGTGACTGGACTCATTTCAGGGAGCTC 1053
Db 61 tggctgcaatattagttgctgtatgtatgcgattcttcctcaagaagaactctcttta 120
QY 1054 TGGTTGCAATATTAGTTGCTGTATGTATGTCGGATTCTTCAAAGAAAGAACTTCTTTTA 1113
Db 121 aagaagaagaagaggagactctcatcaactctgcattgataaacacacacactgggaatc 180
QY 1114 AAGAAAGAAAAGAGGAGGACTCTCATCAACTCTGCATGAACACACCAACACTGGGAATC 1173
Db 181 actatnccagcgcaatcaaccagccttgaaggcagcaggtgccaggtgaagctggcc 240
QY 1174 ACTATC--C-GAGCAATCACCAGCCTTGAAGGCGAGCGGTGCCAGGTGAAGCTGGCC 1230
Db 241 tgttttctaaggaagaatgattgccacaagn-aagagatgcattcttctcctgtgta 299
QY 1231 TGTTTCTAAGGAAATGATTGCCAAGCAAGAGGATGCATCTTCTCTCGGTGTA 1290
Db 300 caagcctttaagactctgctgctgctaatgctcttggatgcacactttgtgtgta 359
QY 1291 CAAGCCTTTAAGACTTCTGCTGTGATA-TGCCCTTGGATGCACACTTTGTGTGTACA 1349
Db 360 tagttaccttaactcagtggtgtatctaataagcttaactcattaaaaaactcc 415
QY 1350 TAGTTAGCTTTAACTCAGTGGTTATCTAATAGCTTAACTCAATAAAAAAACTCC 1405

Search completed: Thu Nov 6 11:26:56 1997
Job time : 983 secs.

This Page Blank (uspto)

WILEY

(TM)

Release 2.1D John F. Collins, BioComputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.
MPsrch_un n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Thu Nov 6 11:27:18 1997; Maspar time 421.77 Seconds
Tabular output not generated.
1020.354 Million cell updates/sec

Title: >US-08-842-827-3
Description: (1-1566) from US08842827.seq
Perfect Score: 1566
N.A. Sequence: 1 CCTGTGGAGAGAGCGCGG.....CCRAAAAAAAAAAAAAA 1566
Comp: GGACACCCCTCTCGCGGCC.....GGTTTTTTTTTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 359085 seqs, 137405154 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: EST-STS-THREE
1:EST199 2:EST200 3:EST201 4:EST202 5:EST203 6:EST204
7:EST205 8:EST206 9:EST207 10:EST208 11:EST209 12:EST210
13:EST211 14:EST212 15:EST213 16:EST214 17:EST215
18:EST216 19:EST217 20:EST218 21:EST219 22:EST220
23:EST221 24:EST222 25:EST223 26:EST224 27:EST225
28:EST226 29:EST227 30:EST228 31:EST229 32:EST230
33:EST231 34:EST232 35:EST233 36:STS1 37:STS2 38:STS3
39:STS4 40:STS5 41:STS6 42:STS7 43:STS8 44:STS9 45:STS10
46:STS11 47:STS12 48:STS13
EST-STS-FOUR
49:gnEST1 50:gnEST2 51:gnEST3 52:gnEST4 53:gnEST5
54:gnEST6 55:gnEST7 56:gnEST8 57:gnEST9 58:gnEST10
59:gnEST11 60:gnEST12 61:gnEST13 62:gnEST14 63:gnEST15
64:gnEST16 65:gnEST17 66:gnEST18 67:gnEST19 68:gnEST20
69:gnEST21 70:gnEST22 71:gnEST23 72:gnEST24 73:gnEST25
74:gnEST5 75:gnEST11 76:gnEST12 77:gnEST13 78:gnEST14
79:gnEST5 80:gnEST6 81:gnEST7 82:gnEST8 83:gnEST9
84:gnEST10 85:gnEST11 86:gnEST12 87:gnEST13 88:gnEST14
89:gnEST15 90:gnEST16 91:gnEST17 92:gnEST18 93:gnEST19
94:gnEST20 95:gnEST21 96:gnEST22 97:gnEST23 98:gnEST24
99:gnEST25 100:gnEST26 101:gnEST27 102:gnEST28
103:gnEST29 104:gnEST30 105:gnEST31 106:gnEST32 107:ueEST1
108:ueEST2

Database:

Statistics: Mean 11.842; Variance 3.114; scale 3.803

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	482	30.8	514	28	AA024599	ze78b09.r1 Soares fet	0.00e+00
2	370	24.3	390	9	AA132640	zo20d08.r1 Stratagene	0.00e+00
3	386	24.0	415	87	HS605336	zb3f08.r1 Soares fet	0.00e+00
4	359	22.9	376	9	AA132534	zo30d08.s1 Stratagene	0.00e+00
5	352	22.5	378	28	AA024600	ze78b09.s1 Soares fet	0.00e+00
6	345	22.0	377	30	AA062861	zf70e10.s1 Soares pin	0.00e+00
7	286	18.3	312	30	AA062853	zf70d09.r1 Soares pin	0.00e+00
8	259	16.5	385	33	AA107326	ml95b10.r1 Stratagene	0.00e+00
9	229	14.6	542	22	AA210789	zr50e08.r1 Soares NH	4.47e-298
10	229	14.6	542	54	AA210789	zr50e08.r1 Soares NH	4.47e-298
11	120	7.7	241	87	HS368337	zc68b11.s1 Soares fet	2.15e-130
12	119	7.6	501	43	G23282	human STS WI-15758	6.68e-129
13	88	5.6	313	91	MM1155742	mx78g01.r1 Soares mou	1.86e-83
14	88	5.6	313	66	AA237956	mx78g01.r1 Soares mou	1.86e-83
15	76	4.9	564	88	HS942332	zc64f09.r1 Soares fet	1.97e-86
16	76	4.9	564	1	W30942	zc64f09.r1 Soares fet	1.97e-86
17	61	3.9	411	50	AA152779	mr78c03.r1 Stratagene	7.62e-46
18	58	3.7	87	33	AA106892	ml85a07.r1 Stratagene	7.54e-42
19	53	3.4	242	27	W39815	308 Mouse VM CDNA lib	2.64e-35
20	53	3.4	337	27	W39816	309 Mouse VM CDNA lib	2.64e-35
21	53	3.4	402	27	W39811	307 Mouse VM CDNA lib	2.64e-35
22	46	2.9	320	10	AA015411	mh21g01.r1 Soares mou	1.96e-26
23	44	2.8	478	26	N75714	yv31b01.r1 Soares fet	5.64e-24
24	42	2.7	185	90	MM0272	mb69e06.r1 Soares mou	1.49e-21
25	40	2.6	184	95	MM63211	mb78e01.r1 Soares mou	3.55e-19
26	40	2.6	199	10	AA014140	mh29b02.r1 Soares mou	3.55e-19
27	40	2.6	269	95	MM5326	mb22b07.r1 Soares mou	3.55e-19
28	37	2.4	52	49	AA122946	mq69e06.r1 Stratagene	3.55e-19
29	37	2.4	52	6	AA122946	mq69e06.r1 Stratagene	3.55e-19
30	34	2.2	369	24	AA217401	mu30f02.r1 Soares mou	1.07e-15
31	34	2.2	429	56	AA144221	mr14h12.r1 Soares mou	2.42e-12
32	33	2.1	236	81	HS1146241	zp52f10.r1 Stratagene	2.95e-11
33	33	2.1	236	12	AA182434	zp52f10.r1 Stratagene	2.95e-11
34	33	2.1	301	56	AA138853	mr03a06.r1 Soares mou	2.95e-11
35	29	1.9	181	67	AA241676	MBL2SJ8B5T3 JHU96SL-B	4.34e-07
36	29	1.9	181	79	BNAA41676	MBL2SJ8B5T3 JHU96SL-B	4.34e-07
37	29	1.9	197	70	AA208192	mv91b08.r1 GuayWoodfo	4.34e-07
38	29	1.9	197	20	AA208192	mv91b08.r1 GuayWoodfo	4.34e-07
39	29	1.9	197	103	MM481921	mv91b08.r1 GuayWoodfo	4.34e-07
40	29	1.9	339	1	N97773	1238C3 czappPFD2.1, D	4.34e-07
41	30	1.9	446	12	AA181403	zp52f10.s1 Stratagene	4.22e-08
42	28	1.8	153	15	AA189532	mt90c05.r1 Soares mou	4.25e-06
43	28	1.8	244	5	AA117061	mn29d08.r1 Beddington	4.25e-06
44	28	1.8	453	105	MMW9541	mf84g07.r1 Soares mou	4.25e-06
45	28	1.8	1097	73	C22771	Dictyostellium discoid	4.25e-06

ALIGNMENTS

RESULT	1	AA024599	514 bp	mrna	EST	01-FEB-1997
LOCUS		ze78b09.r1 Soares fetal heart				
DEFINITION		365081 5'				
ACCESSION		AA024599				
NID		g1489504				
KEYWORDS		EST.				
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		1 (bases 1 to 514)				
AUTHORS		Hallier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Rukhman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.				
TITLE		WashU-Merck EST Project				
JOURNAL		Unpublished (1995)				
COMMENT		Contact: Wilson RK WashU-Merck EST Project Washington University School of Medicine				

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 600 Std Error: 0.00

Seq primer: -28M13 rev2 from AmerSham

High quality sequence stop: 478.

FEATURES

Source
1..514
/organism="Homo sapiens"
/note="Organ: heart; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGCTTACCAATCTGAGTGGGAGCGCGCATCTTTTCTTTTCTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NDHL19W."
/clone="365081"
/clone_lib="Soares fetal heart MbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
<1..>514

BASE COUNT 162 a 98 c 97 g 153 t 4 others
ORIGIN

Query Match 30.8%; Score 482; DB 28; Length 514;
Best Local Similarity 98.4%; Pred. No. 0.00e+00;
Matches 507; Conservative 0; Mismatches 4; Indels 4; Gaps 4;
Db 1 ctggactcttcaggagctctgttgcaatattagttgctgtatatgtatcgattctt 60
QY 1033 CTGGACTCTTCAGGAGCTCTGTTGCAATATTAGTGTGTATATGATCGGATTTCT 1092
Db 61 tcaagaagaacctcttttaagaagaanaagagagagactctctacaactctgcag 120
QY 1093 TCAAGAAAAGAACTCTTTTAAAGAAAAGAAAGAGAGGAGACTCTCATACACTCTGCATG 1152
Db 121 aaacacacaactcgggaatcactatccagcaatcacagccttgaaggcagcagggt 180
QY 1153 AAACACCACTGGGAATCCTATCCGAGCAATCACCAGCCTTGAAGGCAGCAGGGT 1212
Db 181 gccagggtgaagctggcctgttttctaaaggaaaatgattgccacaagn-aagaggatgc 239
QY 1213 GCCCAGGTGAAGCTGCCCTGTTTCTAAAGGAAAATGATTGCCACAAGCAGAGATGC 1272
Db 240 attcttctcctggtgtacaagcctttaagacctctgtcgtcgtatgcctcttgatg 299
QY 1273 ATCTTTCTTCCTGGTGTACAAAGCCTTTAAAGACTTCTGCTGTATGCTCTTGGATG 1332
Db 300 cacacttctgtcatagttaccttcaactcagtggttatctaatagctctaaactca 359
QY 1333 CACACTTTGTGTATAGTTTACCTTTTAACTCAGTGGTGTATCTAATAGCTCTAAACTCA 1392
Db 360 ttaaaaaactccagccttccaccacaaacagtgccccacctgtatcacatttttataaa 419
QY 1393 TTAATAAACTCAAGCCTTCACCAAAACAGTGCCCCACCTGTATACATTTTATATAA 1452
Db 420 aaaaatgaatgcttatggtataaaactgggtatgtaataatgcttctctaatgatgttg 479
QY 1453 AAAATGTAATGCTTATG-TATAAACATG-TATGTAATATGCTTCTCTATGATGATGTTG 1510
Db 480 atttaaatatacatatataataatgattggng 514
QY 1511 ATTTAAAT-TATAATACATTTAAATATGATGGAG 1544

RESULT 2

LOCUS AA132640 390 bp mRNA EST 27-NOV-1996
DEFINITION ZO30d08.r1 Stratagene colon (#937204) Homo sapiens cDNA clone 587439 5'.
ACCESSION AA132640
NID 91694147
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondria eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 390)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.
WashU-Merck EST Project
Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28M13 rev2 from AmerSham.
FEATURES
Location/Qualifiers
1..390
/organism="Homo sapiens"
/note="Organ: colon; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. T-84 colonic epithelial cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3".
/clone="587439"
/clone_lib="Stratagene colon (#937204)"
/lab_host="SOLR cells (kanamycin resistant)"
<1..>390
BASE COUNT 111 a 81 c 82 g 115 t 1 others
ORIGIN
Query Match 24.3%; Score 380; DB 9; Length 390;
Best Local Similarity 99.2%; Pred. No. 0.00e+00;
Matches 387; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
Db 1 gaaggcttgagcttttttaagagatttagagctattagataaacacaggttaaaggta 60
CP 1413 GAAGGCTTGGAGTTTTTTTAAATGAGTTAGAGCTATTAGATAACCACCTGAGTTAAGGTA 1354
Db 61 actatgtacacacaagtggtgcaatcccaaggagcctagcagcagcagagcttttaaagg 120
CP 1353 ACTATGTACACAAAGTGTGCA-TCCAAAGAGCATATCAGCAGCAGAGAGTCTTTAAAGG 1295
Db 121 ctgtacaccaggaagaagatgcatctcttgccttggcgaatcatttccctcttagaa 180
CP 1294 CTTGTACACAGGAAGAAAGATGTCATCTCTTGCCTTGTGGCAATCATTTCTCTTAGAA 1235
Db 181 aacagggcagcttcacctggggcacctctgctcttcaagnotggtgattgctcgatag 240
CP 1234 AACAGGCCAGCTTCACCTGGGCACCTCTGCTCTTTCAGGCTGGTATTGTCGGATAG 1175
Db 241 tgattccagctgttgggtttccatgcagagttgtatgagagctcctctcttcttcttct 300
CP 1174 TGATTCACAGTTGTTGGGTGTTTCATGCGAGAGTTGTATGAGAGTCTCTCTCTTTCTTCT 1115
Db 301 ttaagaaggtcttcttcttgaagaaatccgatacatatcacagcaactaatattgcaacc 360
CP 1114 TTAAGAAGGTTCTTTCTTTGAAGAAATCCGATACATATACAGCAACTAATATTGAACC 1055

**JOURNAL
COMMENT**

Unpublished (1996)

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

FEATURES

```

/organism="Mus musculus"
/strain="C57/BL6"
/note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2:
XhoI; Cloned unidirectionally. Primer: Oligo dT. Average
insert size: 1.0 kb; Uni-zap XR vector; -5' adaptor
sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5'
CTCGAGCTTTTTTTTTTTTTT 3'"
/clone="519739"
/clone_lib="Stratagene mouse kidney (#937315)"
/sex="females"
/dev_stage="4 weeks"
/lab_host="SOLR (kanamycin resistant)"

```

RESULT	9	542 bp	mRNA	EST	31-JAN-1997
LOCUS	AA210789				
DEFINITION	zr9e08.r1 Soares NbHTGBC Homo sapiens				
ACCESSION	AA210789				
NID	91809443				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				

REFERENCE
AUTHORS

TITLE	JOURNAL	COMMENT

Contact: Wilson RK
WashU-Merk EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LINL & con-
IMAGE Consortium (info@image.lnl.gov) for further info
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 414

mrna	BASE COUNT	ORIGIN
1	100	1
2	100	2
3	100	3
4	100	4
5	100	5
6	100	6
7	100	7
8	100	8
9	100	9
10	100	10
11	100	11
12	100	12
13	100	13
14	100	14
15	100	15
16	100	16
17	100	17
18	100	18
19	100	19
20	100	20
21	100	21
22	100	22
23	100	23
24	100	24
25	100	25
26	100	26
27	100	27
28	100	28
29	100	29
30	100	30
31	100	31
32	100	32
33	100	33
34	100	34
35	100	35
36	100	36
37	100	37
38	100	38
39	100	39
40	100	40
41	100	41
42	100	42
43	100	43
44	100	44
45	100	45
46	100	46
47	100	47
48	100	48
49	100	49
50	100	50
51	100	51
52	100	52
53	100	53
54	100	54
55	100	55
56	100	56
57	100	57
58	100	58
59	100	59
60	100	60
61	100	61
62	100	62
63	100	63
64	100	64
65	100	65
66	100	66
67	100	67
68	100	68
69	100	69
70	100	70
71	100	71
72	100	72
73	100	73
74	100	74
75	100	75
76	100	76
77	100	77
78	100	78
79	100	79
80	100	80
81	100	81
82	100	82
83	100	83
84	100	84
85	100	85
86	100	86
87	100	87
88	100	88
89	100	89
90	100	90
91	100	91
92	100	92
93	100	93
94	100	94
95	100	95
96	100	96
97	100	97
98	100	98
99	100	99
100	100	100

Query Match 14.6%; Score 229; DB 22; Length 542;
Best Local Similarity 95.4%; Pred. No. 4.47e-298;
Matches 308; Conservative 0; Mismatches 2; Indels 1

RESULT	10
--------	----


```

RESULT 12
LOCUS G23282 501 bp DNA STS 31-MAY-1996
DEFINITION human STS WI-15758.
ACCESSION G23282
NID 91343608
KEYWORDS STS sequence; primer; sequence tagged site.
SOURCE human STSs derived from sequences in dbEST and the Unigene
collection.

ORGANISM Homo sapiens
Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 501)
AUTHORS Hudson,T.
TITLE Whitehead Institute/MIT Center for Genome Research; Physically
Mapped STSs
JOURNAL Unpublished (1995)
COMMENT
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu

Primer A: TATTCAGTATTGGCAAAATGAAGC
Primer B: TACTTTGAGGTGTACAGTTAAGCA
STS size: 129
PCR Profile:
  Presoak:
  Denaturation:
  Annealing: 56 degrees C
  Polymerization:
  PCR Cycles: 35
  Thermal Cycler:
Protocol:
  Template: 10 ng
  Primer: each 5 pM
  dNTPs: each 4 nM
  Taq Polymerase: 0.025 units/ul
  Total Vol: 20 ul

Buffer:
  MgCl2: 1.5 mM
  KCl: 50 mM
  Tris-HCl: 10 mM
  pH: 9.3

Derived from dbEST (genbank accession R45479).
Location/Qualifiers
  1..501
/organism="Homo sapiens"
/map="188.1 CR from top of Chr5 linkage group"
23..151
primer_bind 23..46
primer_bind complement(127..151)
BASE COUNT 133 a 88 c 117 g 157 t 6 others
ORIGIN

Query Match 7.6%; Score 119; DB 43; Length 501;
Best Local Similarity 84.9%; Pred. No. 6.68e-129;
Matches 202; Conservative 0; Mismatches 26; Indels 10; Gaps 10;

Db 255 gctgtatagtctcgtggtttcttcaagaaggaagactcttttaagaagaagaagagg 314
|||||
QY 1071 GCTGTATATGTATCGGATTCTCTCAAGAAGAAGAACTCTTTTAAAG-AAAG-AAAGAGG 1128
|||||

Db 315 ggggggtctcatcaactctggcgtggaacacccnacaactggggttccactatccgagg 374
|||||
QY 1129 AGGA-CTCTCATACACTCTG-CATG-AAACACCAACAACACTGGGAAT-CACATCCGAG- 1183
|||||

```

```

Db 375 caatcaccagccttgaagggcgcggggtgccgggggtgagggcggggctgtttcttaa 434
|||||
QY 1184 CAATCACCAGCCTTGAAGG-CAGCAGGGTGCCAGG-TGAAGCTGG-CCTGTTTCTTAA 1240
|||||

Db 435 ggggaaatggtgtccacaggnccagggggtgctcttcttcttcctccgggggtacaggttt 492
|||||
QY 1241 AGGAAATGATTGCCACAGGCAAGAGGATGCATCTTCTCTCTGCTGTACAAAGCCTT 1298
|||||

RESULT 13
ID M01155742 standard; RNA; EST; 313 BP.
AC AA237956;
NI G1862038
DT 06-MAR-1997 (Rel. 51, Created)
DT 06-MAR-1997 (Rel. 51, Last updated, Version 1)
DE mx76g01.r1 Soares mouse NML Mus musculus cDNA clone 692496 5'
DE similar to TR:G1161100 G1161100 HYDROGEN PEROXIDE-INDUCIBLE PROTEIN
DE ;
KW EST.
OS Mus musculus (house mouse)
OC Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
OC Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
OC Mus.
RN [1]
RP 1-313
RA Marra M., Hillier L., Allen M., Bowles M., Dietrich N.,
RA Dubuque T., Geisel S., Kucaba T., Lacy M., Le M., Martin J.,
RA Morris M., Schellenberg K., Steptoe M., Tan F., Underwood K.,
RA Moore B., Theising B., Wylie T., Lennon G., Soares B., Wilson R.,
RA Waterston R.;
RT "The WashU-HHMI Mouse EST Project";
RL Unpublished.
CC Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project
CC Washington University School of Medicine 444 Forest Park Parkway,
CC Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810
CC Email: mouseest@wustl.edu This clone is available
CC royalty-free through LNL; contact the IMAGE Consortium
CC (info@image.llnl.gov) for further information. MGI:426056 Possible
CC reversed clone: similarity on wrong strand Seq primer: 28m13 rev2
CC ET from Amersham High quality sequence stop: 57.
FH Key
PH Location/Qualifiers
FT source
FT 1..313
FT /organism="Mus musculus"
FT /note="vector: pT73D-Pac (Pharmacia) with a modified
FT polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
FT was primed with a Not I - oligo(dT) primer [5',
FT TGTTACCAATCTGAAGTGGGCGCGCGCAATCTTTTTTTTTTTT 3'];
FT double-stranded cDNA was ligated to Eco RI adaptors
FT (Pharmacia), digested with Not I and cloned into the Not I
FT and Eco RI sites of the modified pT73 vector. Library
FT constructed and normalized by Bento Soares and M.Fatima
FT Bonaldo."
FT /clone="692496"
FT /clone_lib="Soares mouse NML"
FT /tissue_type="Liver"
FT /lab_host="DH10B"
FT <1..>313
FT MRNA
SQ Sequence 313 BP; 83 A; 58 C; 85 G; 87 T; 0 other;

Query Match 5.6%; Score 88; DB 91; Length 313;
Best Local Similarity 66.5%; Pred. No. 1.86e-83;
Matches 208; Conservative 0; Mismatches 102; Indels 3; Gaps 3;

Db 3 gaggccaactgtgagctcactccagtggtgtttgttagtcagacactcgaagaagcc-ac 61
|||||
Cp 1040 GAGTCAGTCACACATCGCTCCAGTGTGTTTATATCAGAAACTCGAAGAGCCAC 981
|||||

Db 62 atatataaataaagcaatgagccaaactggagcaatgtctgaagaagcctggccaatt 121
|||||
QY 980 ATAAATGATACGCGCAACAGACCAATTCAGTGTGGGGCGTAAGA-GTCTTGCCCACT 922
|||||

Db 122 ttcccttaactcgtggcttgaggttaagttgcgaacacaggattgagttacttgaggattg 181
|||||

```

```

Cp 921 CTCCTTCATCTCGTGTGAAGATAAAGTCCACAAACAGCATGCAGTACATGGAACG 862
    ||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 agaatttcgggttggaagccaactggcctctcttgatttccctt-caatccctggc 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 861 AAGAGTGGCCTGAATAGAGCAACCTGCTTCCCTTAACCTCTTTCGCAATCCCTCGAC 802
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 ataagsgttcttcattgcatgaatgaagttggatttgaacagctctcggttataga 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 801 ATATGTAGTATTCAATGATACCATCGCTGCAGTTGATTTTGACCAATCTGGATCACA 742
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 gatgccagaagt 313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 741 CATCCAAGAAGTG 729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 14
LOCUS AA237956 313 bp mRNA EST 03-MAR-1997
DEFINITION mx78901.r1 Soares mouse NML Mus musculus cDNA clone 692496 5'
similar to TR:G1161100 G1161100 HYDROGEN PEROXIDE-INDUCIBLE PROTEIN
;

```

```

ACCESSION AA237956
NID G1862038
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Mus.

```

```

REFERENCE 1 (bases 1 to 313)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT

```

Contact: Marra M/Mouse EST Project
WashU-HMMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:426036

Possible reversed clone: similarity on wrong strand

Seq primer: -28m13 rev2 ET from Amersham

High quality sequence stop: 57.

Location/Qualifiers

source

```

1..313
/organism="Mus musculus"
/notes="vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGCGCGCGCAATCTTTTCTTTTCTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. Library
constructed and normalized by Bento Soares and M.Fatima
Ronald."
/clone="692496"
/clone_lib="Soares mouse NML"
/tissue_type="Liver"
/lab_host="DH10B"
<1..>313

```

```

BASE COUNT 83 a 58 c 85 g 87 t
ORIGIN

```

Query Match 5.6%; Score 88; DB 66; Length 313;
Best Local Similarity 66.5%; Pred. No. 1.86e-83;

```

Matches 208; Conservative 0; Mismatches 102; Indels 3; Gaps 3;
Db 3 gagtccaactgtgacgtcactccagtggtttagtcagacacacgcgaaaggcc-ac 61
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 1040 GAGTCAGTCAACACATCGCTCCAGTGGTGTATATATCAGAAATCTCGAAAGGCCAC 981
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 atatatgaaaaagcaatgagggccaaactggagcaatggtcgttaagaagcctggccaatt 121
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 980 ATAAATGGATACGGCAACAGACAAATTCAGTGTGGGGCGTAAGA-GTCTTGCACGT 922
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 ttccttaatactgggttgaggttaagttgacacacagagattgagttacttggggattg 181
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 921 CTCCTTCATCTCGTGTGAAGATAAAGTCCACAAACAGCATGCAGTACATGGAACG 862
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 agaatttcgggttggaagccaactggcctctctggatttccctt-caatccctggc 240
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 861 AAGAGTGGCCTGAATAGAGGACAACCTGCTTCCCTTAACCTCTTTCGCAATCCCTCGAC 802
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 ataagsgatttcttcattgcatgaatgaagttggatttgaacagctctcggttataga 300
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 801 ATATGTAGTATTCAATGATACCATCGCTGCAGTTGATTTTGACCAATCTGGATCACA 742
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 gatgccagaagt 313
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 741 CATCCAAGAAGTG 729
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 15
ID HS942332 standard; RNA; EST; 564 BP.
AC W30942;
NI G1311934
DT 13-MAY-1996 (Rel. 47, Created)
DT 07-MAR-1997 (Rel. 51, Last updated, Version 2)
DE z664f09.r1 Soares fetal heart NBHH19W Homo sapiens cDNA clone
DE 327113 5' similar to WP:728D9.3 CE02068 ;
KW EST.
OS Homo sapiens (human)
OC Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
OC Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP 1-564
RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,
RA Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,
RA Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F.,
RA Trevisakis E., Waterston R., Williamson A., Wohlmann P., Wilson R.;
RA "The WashU-Merck EST Project";
RL Unpublished.
CC Contact: Wilson RK WashU-Merck EST Project Washington University
CC School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,
CC MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:
CC est@wustl.edu This clone is available royalty-free through
CC LNL; contact the IMAGE Consortium (info@image.llnl.gov) for
CC further information. Insert length: 774 Std Error: 0.00 Seq primer:
CC mob.REGA+ET High quality sequence stop: 336.
Key Location/Qualifiers

```

```

source 1..564
/organism="Homo sapiens"
/notes="Organ: heart; Vector: p7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGCGCGCGCAATCTTTTCTTTTCTTTT 3'];
double-stranded cDNA was size selected, ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by M.Fatima
Ronald."
Bonaldo. This library was constructed from the same fetus
as the fetal lung library, Soares fetal lung NBHH19W."
/clone="327113"
/clone_lib="Soares fetal heart NBHH19W"
/sex="unknown"

```

```
FT /dev_stage="19 weeks"
FT /lab_host="DH10B (ampicillin resistant)"
FT mRNA <1..>564
SQ Sequence 564 BP; 128 A; 155 C; 139 G; 134 T; 8 other;

Query Match 4.9%; Score 76; DB 88; Length 564;
Best Local Similarity 67.8%; Pred. No. 1.97e-66;
Matches 158; Conservative 0; Mismatches 73; Indels 2; Gaps 2;

Db 251 agccagctcttcacagacattgccaaagtgtccatagggcgctggtcctcacttcttg 310
QY || ||||| || ||||| ||||| || ||||| ||||| ||||| ||||| |||||
Db 678 AGTCAGTCCCTGACTGACATTGCCAAGTATTCAATAGGCAGACTGCGGCTCACTTCTTG 737
QY || ||||| || ||||| ||||| || ||||| ||||| ||||| ||||| |||||
Db 311 aagtgtctgcaacccttgattccagccagatccaactgctctgaagggtacattccagaacta 370
QY || || || || || || || || || || || || || || || || || || || || ||
Db 738 GA-TGTTTGATCCAGATTGGTCAAAATCAACTGCGGATGGTTACATTCAATACTA 796
QY || || || || || || || || || || || || || || || || || || || || ||
Db 371 cagatgcagaggtgatgacagcaaaagtcaggangccagggaagtccttctcttgcca 430
QY || || || || || || || || || || || || || || || || || || || || ||
Db 797 CATATGTCGAGGGAATGCAGAAAGAGTTAAGGAAGCAGGTTGTCCTTCTATTTCAGGCCA 856
QY || || || || || || || || || || || || || || || || || || || || ||
Db 431 tgcctcttctccatgacactatgctgaatttggtgctatancctgcaggcc 483
QY || || || || || || || || || || || || || || || || || || || || ||
Db 857 CTCCTGCTTTTCATGTACTGATGCTGTT-TGTGGCACTTTATCTTCAAGCC 908
QY || || || || || || || || || || || || || || || || || || || || ||
```

Search completed: Thu Nov 6 11:35:37 1997
Job time : 499 secs.

WPSRL

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Nov 4 10:23:32 1997; Maspar time 7.03 Seconds
Tabular output not generated.
498.122 Million cell updates/sec

Title: >US-08-842-827-4
Description: (1-285) from US08842827.pap
Perfect Score: 2119
Sequence: 1 MFDKTRLPYVALDVLCLIA.....HTLHETPTGNHPSNHQP 285

Scoring table: PAM 150
Gap 11

Searched: 101610 seqs, 12294212 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq28
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21

Statistics: Mean 33.649; Variance 149.045; scale 0.226

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	109	5.1	193	21	W20304 H. pylori cytoplasmic	1.86e+00
2	109	5.1	229	21	W20620 H. pylori cytoplasmic	1.86e+00
3	105	5.0	682	19	W04359 S. pneumoniae penicil	3.63e+00
4	96	4.5	397	18	W04217 Rat dorsal root gangl	1.59e+01
5	93	4.4	317	19	W02661 G-protein coupled hum	2.56e+01
6	93	4.4	317	16	W48689 G-protein coupled hum	2.56e+01
7	94	4.4	418	7	R39263 Human somatostatin re	2.18e+01
8	94	4.4	563	10	R53463 Pea proton transport	2.18e+01
9	94	4.4	1073	10	R55272 Alpha subunit of inte	2.18e+01
10	94	4.4	1073	5	R38821 Alpha 6A integrin sub	2.18e+01
11	92	4.3	276	21	W20612 H. pylori protein.	3.00e+01
12	91	4.3	1091	5	R38822 Alpha 6B integrin sub	3.51e+01
13	92	4.3	1539	21	W33905 Human SMCY protein.	3.00e+01
14	91	4.3	3011	7	R34468 Encoded by full-length	3.51e+01
15	88	4.2	170	21	W20704 H. pylori transmembra	5.62e+01
16	90	4.2	205	21	W20268 H. pylori transporter	4.11e+01
17	89	4.2	250	6	R38365 Human LFA-3 protein.	4.81e+01
18	90	4.2	270	21	W20935 H. pylori transporter	4.11e+01
19	90	4.2	470	7	R34467 Encoded by Hepatitis	4.11e+01
20	90	4.2	470	7	R34476 Encoded by Hepatitis	4.11e+01

21	90	4.2	470	7	R34478	Encoded by Hepatitis	4.11e+01
22	90	4.2	496	21	W20801	H. pylori inner membr	4.11e+01
23	88	4.2	628	5	R27575	ABF-A from A. niger.	5.62e+01
24	88	4.2	758	6	R28954	Vitamin K dependent c	5.62e+01
25	88	4.2	1063	13	R73916	Rubella virus strain	5.62e+01
26	88	4.2	1141	6	R31961	Human cardiac cgl PDE	5.62e+01
27	90	4.2	1189	11	R56496	TATA-binding protein-	4.11e+01
28	88	4.2	1353	19	R99251	Murine adenylate cycl	5.62e+01
29	90	4.2	1463	12	R63575	Group 1 phospholipase	4.11e+01
30	90	4.2	1463	15	R85143	Bovine phospholipase	4.11e+01
31	86	4.1	270	18	R99586	Low density lipoprote	7.66e+01
32	86	4.1	273	18	R99587	Low density lipoprote	7.66e+01
33	87	4.1	289	12	R67996	Novel mouse proteogly	6.56e+01
34	87	4.1	310	17	R87000	Human syndecan-1.	6.56e+01
35	87	4.1	311	12	R66793	Novel mouse proteogly	6.56e+01
36	87	4.1	311	10	R55276	Syndecan protein.	6.56e+01
37	87	4.1	311	17	R87001	Mouse syndecan-1.	6.56e+01
38	87	4.1	389	3	R15428	3-acylation enzyme.	6.56e+01
39	87	4.1	428	7	R39264	Murine somatostatin r	6.56e+01
40	87	4.1	643	18	R97735	Bacillus thuringiens	6.56e+01
41	87	4.1	643	1	R91462	67-kD protein toxin.	6.56e+01
42	87	4.1	1063	15	R79049	Infectious rubella vi	6.56e+01
43	87	4.1	1358	19	R99358	Human restrictin.	6.56e+01
44	86	4.1	1876	14	R72853	S. cerevisiae FKSl.	7.66e+01
45	85	4.0	262	8	R42502	A. rhizogenes NIAES17	8.93e+01

ALIGNMENTS

RESULT 1
ID W20304 standard; Protein; 193 AA.
AC W20304;
DE 10-JUL-1997 (first entry)
DE H. pylori cytoplasmic protein 24407533.aa.
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
OS Helicobacter pylori.
FH Key Location/Qualifiers
FT Misc_difference 100
FT /note="Encoded by GAR"
PN WO9640893-AL.
PD 19-DEC-1996.
PF 06-JUN-1996; U09123.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR) ASTRA AB.
PI Berglindh OT, Smith D, Mellgaard BL;
DR WPI; 97-052306/05.
DR N-PSDB; T67510.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 61; Page 502; 1481pp; English.
CC The present sequence shows a Helicobacter pylori cytoplasmic protein
CC that may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679), was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
SQ Sequence 193 AA;

Query Match 5.1%; Score 109; DB 21; Length 193;
Best Local Similarity 20.8%; Pred. No. 1.86e+00;
Matches 26; Conservative 36; Mismatches 59; Indels 4; Gaps 4;

Db 14 ifdshlhkvegfkalfqkghndskglkfeyhygsgisrnekikyfyneilktpiaq 73
 QY 1 MFDKRLPYVALDVLCVLLASMPMAVLKLGQIYFFQGFCKDINSINYPHSTAASTVL 60
 Db 74 eedlalefagai-eqklfdgrhlsevmafidkhyqnyfihiaaahselqlvicefl 132
 QY 61 ILV-GVGLPVSSIIIGETLSVYCNLL-HSNSFISNNYIATIKYKAIGTFLFGAAASQS-LT 117
 Db 133 gicky 137
 QY 118 DIAKY 122

RESULT 2

ID W20620 standard; Protein: 229 AA.
 AC W20620;
 DT 14-JUL-1997 (first entry)
 DE H. pylori cytoplasmic protein, 02aell161lorfs.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacterium; life cycle; activator;
 KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
 KW diagnosis.
 OS Helicobacter pylori.
 PN W09640893-A1.
 PD 19-DEC-1996.
 PF 06-JUN-1996; U09122.
 PR 07-JUN-1995; US-487032.
 PR 01-APR-1996; US-630405.
 PA (ASTR) ASTRA AB.
 PI Berglindh OT, Smith D, Meilgaard BL;
 DR WPI: 97-052306/05.
 DR N-PSDB: T67873.
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptides) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 PS Claim 61; Page 1047; 1481pp; English.
 CC This sequence represents a H. pylori cytoplasmic protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed, the amino
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 SQ Sequence 229 AA;

Query Match 5.1%; Score 109; DB 21; Length 229;
 Best Local Similarity 20.8%; Pred. No. 1.86e+00;
 Matches 26; Conservative 36; Mismatches 59; Indels 4; Gaps 4;

Db 21 ifdshlhkvegfkalfqkghndskglkfeyhygsgisrnekikyfyneilktpiaq 80
 QY 1 MFDKRLPYVALDVLCVLLASMPMAVLKLGQIYFFQGFCKDINSINYPHSTAASTVL 60
 Db 81 eedlalefagai-eqklfdgrhlsevmafidkhyqnyfihiaaahselqlvicefl 139
 QY 61 ILV-GVGLPVSSIIIGETLSVYCNLL-HSNSFISNNYIATIKYKAIGTFLFGAAASQS-LT 117
 Db 140 gicky 144
 QY 118 DIAKY 122

RESULT 3

ID W04359 standard; Protein: 682 AA.
 AC W04359;

DT 03-DEC-1996 (first entry)
 DE S. pneumoniae penicillin binding protein 1A soluble variant.
 KW Penicillin binding protein; PBP 1A; bifunctional protein; inhibitor;
 KW transglycosylase; transpeptidase; identification; assay; protein structure;
 KW antibiotic resistant; bacteria; soluble variant; protein structure;
 KW X-ray crystallography; determination.
 OS Streptococcus pneumoniae.
 PN GB2290792-A.
 PD 10-JAN-1996.
 PF 29-JUN-1995; 013306.
 PR 01-JUL-1994; IN-000580.
 PR 24-NOV-1994; SE-004072.
 PA (ASTR) ASTRA AB.
 PI Balganeshts Town CM;
 DR WPI: 96-042232/05.
 DR N-PSDB: T08027.
 PT opt. derivs. of bifunctional penicillin binding protein (BPPP) -
 PT sol. lack transglycosylase activity, useful to identify and assay
 PT for antibodies or cpds. which bind BPPs
 PS Claim 2; Pages 65-68; 108pp; English.
 CC The present sequence is a soluble variant of the S. pneumoniae
 CC penicillin binding protein (PBP) 1A, which lacks the 1st 38 amino
 CC acid residues of the wild type protein. Wild type PBP is a
 CC bifunctional protein, which binds the cell membrane when expressed
 CC in a bacterial cell, having transglycosylase and transpeptidase
 CC activities. The variant protein (NCIMB 40665) in conjunction with
 CC a labelled anti-bifunctional PBP monoclonal antibody, can be used
 CC to identify and assay for cpds. which bind bifunctional PBP. Such
 CC cpds., as inhibitors of bifunctional PBP have a potential use in
 CC therapeutic cpds. which inhibit the growth of antibiotic resistant
 CC bacteria. The soluble variant may also be used in X-ray
 CC crystallography.
 SQ Sequence 682 AA;

Query Match 5.0%; Score 105; DB 19; Length 682;
 Best Local Similarity 28.6%; Pred. No. 3.63e+00;
 Matches 18; Conservative 20; Mismatches 22; Indels 3; Gaps 3;

Db 371 vynydrgrf-gnithlqalqgrnvpvetlnkvlnrakfl-ngldgypshysnal 428
 QY 32 IYFFQGFCKDINSINYPHSTAASTVLILVGLVPSIIIGETLSV-YCNLLHNSF 90
 Db 429 ssn 431
 QY 91 ISN 93

RESULT 4

ID W04217 standard; Protein: 397 AA.
 AC W04217;
 DT 23-NOV-1996 (first entry)
 DE Rat dorsal root ganglion P2x receptor.
 KW ATP P2x receptor; purinoceptor; ligand-gated ion channel; agonist;
 KW antagonist; epilepsy; cognition; emesis; pain; asthma;
 KW peripheral vascular disease; hypertension; irritable bowel syndrome;
 KW premature ejaculation; cystitis; therapy.
 OS Rattus sp.
 PN W09533048-A2.
 PD 07-DEC-1995.
 PF 24-MAY-1995; E01968.
 PR 27-MAY-1994; GB-010664.
 PR 09-FEB-1995; GB-002480.
 PA (GLAX) GLAXO GROUP LTD.
 PI Buell GN, Valera S;
 DR WPI: 96-030561/03.
 DR N-PSDB: T33854.
 PT DNA encoding ATP P2x receptors of the purinoceptor family - for
 PT screening cpds. useful in treating epilepsy, cognition, emesis,
 PT pain, asthma, peripheral vascular disease, hypertension, etc.
 PS Claim 1; Fig 3; 82pp; English.
 CC The amino acid sequence (W04217) of rat dorsal root ganglion P2x
 CC receptor was deduced from a cDNA clone (T33854) isolated from a
 CC cDNA library. P2x receptors (see also W04215-16 and W04218) are

RESULT 6

ID	R48699	standard; Protein; 317 AA.
AC	R48699;	
DE	05-JUN-1996	(first entry)
DT	G-protein coupled human m2 muscarinic acetylcholine receptor protein.	
DE	G-protein coupled receptor; ligand binding assay; transmembrane domain;	
KW	psychotic disorder; schizophrenia; dopamine; cAMP; adenosine; thrombin;	
KW	muscarinic acetylcholine; adrenergic; endothelin; bombesin; endocrine;	
KW	rhodopsin; opsin; odorant; cytomegalovirus.	
OS	Homo sapiens.	
PN	WO9405695-A1.	
PD	17-MAR-1994.	

```

Db      35  pspaglavsgvlipl-vylvvcvgllgnsliyvvlrhtaspsvtnvylmlaladelf 93
      | : ||| : | : : ||| : | : : | : | : | : | : | : | : | : |
Qy      49  PYHDSTAATVLLVGVGLPVSSI-ILGETLSVYCNLLHSNS-FISNNYIATYKAIGTF 106

```


FT /note- "Putative cation binding domain"
 FT Binding_site 441..449
 FT /note- "Putative cation binding domain"
 FT Domain 1040..1044
 FT /label- Cytoplasmic domain
 FT /note- "Conserved in virtually all integrins"
 PD W09219647-A.
 PN 12-NOV-1992.
 PR 27-APR-1992; U03527.
 PR 03-MAY-1991; US-695564.
 PA (SCRI) SCRIPPS RES INST.
 PI Quaranta V, Tamura RN;
 DR WPI; 92-398799/48.
 DR N-PSDB; Q31188.
 DR Integrin alpha sub-unit cytoplasmic domain polypeptide(s) - used
 PT for prodn. of antibodies and in detection of integrin sub-units
 PT in body samples
 PS Disclosure; Page 69-73; 115pp; English.
 CC The sequences given in R28821-22 are the human alpha 6A and 6B
 CC integrin subunits. Integrins are a family of cell surface receptors
 CC which serve cellular adhesion functions. These receptors form a link
 CC between the extracellular matrix and the cytoskeleton through their
 CC binding to various extracellular components. Each integrin receptor
 CC is a heterodimer comprised of an alpha and a beta subunit. Each alpha
 CC subunit tends to associate with only one type of beta subunit but
 CC there are several exceptions to this rule. These integrins correspond
 CC to the laminin receptor. The cytoplasmic domain of the 6A and 6B
 CC integrins differs from previously isolated alpha 6 integrins. The
 CC human alpha 6B was isolated from human choriocarcinoma cell line JAR
 CC by immunoprecipitation studies.
 CC Sequence 1073 AA;
 SQ

Query Match 4.4%; Score 94; DB 5; Length 1073;
 Best Local Similarity 25.3%; Pred. No. 2.18e+01;
 Matches 20; Conservative 23; Mismatches 29; Indels 7; Gaps 7;
 Db 997 vrvtfpskt-vagysgvvwiilvailagilmallvllwkcggfk-rnk-kdhydat 1053
 Qy 208 VASIIYGLSRSDYK-HHWSDLTGLIQGALV-AILVAV-YVSDFFKRTSFKERKEED 264
 Db 1054 ykhaeihaqpsdkertsd 1072
 Qy 265 SHTT-LHETPTGNHYPN 282

RESULT 11
 ID W20612 standard; protein; 276 AA.
 AC W20612;
 DT 14-JUL-1997 (first entry)
 DE H. pylori protein.
 KW Vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
 OS Helicobacter pylori.
 PN W09640893-A1.
 PD 19-DEC-1996.
 PR 06-JUN-1996; U09122.
 PR 07-JUN-1995; US-487032.
 PR 01-APR-1996; US-630405.
 PA (ASTR) ASTRA AB.
 PI Berglindh OT, Smith D, Melligaard BL;
 DR WPI; 97-052306/05.
 DR N-PSDB; T67865.
 DR Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 FT infection, and to detect Helicobacter
 PS Claim 61; Page 1040-1041; 1481pp; English.
 CC The present sequence is a Helicobacter pylori protein.
 CC The protein may be used in a vaccine to prevent or treat
 CC H. pylori infection or to identify H. pylori polypeptide binding
 CC compounds, useful as potential H. pylori life cycle activators or
 CC inhibitors. The genomic sequence of H. pylori (ATCC 55679) was
 CC determined from overlapping contigs generated by mechanically

CC shearing the bacterial DNA. The sequences were analysed for ORF of
 CC at least 180 nucleotides, and the predicted coding regions defined
 CC by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from
 CC various ORF were analysed for significant homology to other known
 CC or exported membrane proteins. Having identified and determined
 CC the sequences of interest, particular regions can be isolated from
 CC H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 SQ Sequence 276 AA;
 Query Match 4.3%; Score 92; DB 21; Length 276;
 Best Local Similarity 24.7%; Pred. No. 3.00e+01;
 Matches 18; Conservative 21; Mismatches 28; Indels 6; Gaps 6;
 Db 179 emlvsnppiyarnypleksvkephhealfgvgkdeilkelv-flaaklkpfll-vcemg 236
 Qy 83 NLHLSNS-FISNNY-IA-TIYKAIGTFLFGAA-ASQSLTDIAKYSIGRLRPHFLDVCDDP 138
 Db 237 ydqlkslkeclef 249
 Qy 139 WSKINCSDGVIIEY 151

RESULT 12
 ID R28822 standard; Protein; 1091 AA.
 AC R28822;
 DT 23-MAR-1993 (first entry)
 DE Alpha 6B integrin subunit.
 KW Human; alpha 6B; alpha 6B; Integrin; cell surface receptor; adhesion;
 KW extracellular matrix; cytoskeleton; heterodimer; laminin receptor;
 KW immunoprecipitation; JAR; choriocarcinoma.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Domain 1012..1037
 FT /label- Transmembrane domain
 FT Peptide 1..23
 FT /label- Signal peptide
 FT Protein 24..1091
 FT /label- Mature protein
 FT Modified_site 223
 FT /label- Glycosylation_site
 FT Modified_site 284
 FT /label- Glycosylation_site
 FT Modified_site 370
 FT /label- Glycosylation_site
 FT Modified_site 513
 FT /label- Glycosylation_site
 FT Modified_site 731
 FT /label- Glycosylation_site
 FT Modified_site 748
 FT /label- Glycosylation_site
 FT Modified_site 891
 FT /label- Glycosylation_site
 FT Modified_site 927
 FT /label- Glycosylation_site
 FT Modified_site 958
 FT /label- Glycosylation_site
 FT Binding_site 230..238
 FT /note- "Putative cation binding domain"
 FT Binding_site 324..332
 FT /note- "Putative cation binding domain"
 FT Binding_site 385..394
 FT /note- "Putative cation binding domain"
 FT Binding_site 441..449
 FT /note- "Putative cation binding domain"
 FT Domain 1040..1044
 FT /label- Cytoplasmic domain
 FT /note- "Conserved in virtually all integrins"
 FT Misc_difference 1044..1045
 FT /note- "Position of Deletion cf alpha 6A"
 PN W09219647-A.
 PD 12-NOV-1992.

PF 27-APR-1992; U03527.
 PR 03-MAY-1991; US-695564.
 PA (SRI) SCRIPPS RES INST.
 PI Quaranta V, Tamura RN,
 DR WPI: 92-398799/48.
 DR N-PSDB; Q31189.

PT Integrin alpha sub-unit cytoplasmic domain polypeptide(s) - used
 for prodn. of antibodies and in detection of integrin sub-units
 in body samples
 PS Disclosure: Page 78-82; 115pp; English.
 CC The sequences given in R28821-22 are the human alpha 6A and 6B
 integrin subunits. Integrins are a family of cell surface receptors
 which serve cellular adhesion functions. These receptors form a link
 between the extracellular matrix and the cytoskeleton through their
 binding to various extracellular components. Each integrin receptor
 is a heterodimer comprised of an alpha and a beta subunit. Each alpha
 subunit tends to associate with only one type of beta subunit but
 there are several exceptions to this rule. These integrins correspond
 to the laminin receptor. The cytoplasmic domain of the 6A and 6B
 integrins differs from previously isolated alpha 6 integrins. The
 human alpha 6B was isolated from human choriocarcinoma cell line JAR
 by immunoprecipitation studies.
 CC Sequence 1091 AA;

Query Match 4.3%; Score 91; DB 5; Length 1091;
 Best Local Similarity 27.3%; Pred. No. 3.51e+01;
 Matches 15; Conservative 17; Mismatches 18; Indels 5; Gaps 5;

Db 997 vrvtfpskt-vagysgvpwllvalleglmlallvflwxcgffk-rsrydd 1049
 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 208 VAVSIYVGLSRVSDYK-HHWSDLTGLTQALV-AILVAV-YVSDFFKERTSFE 259

RESULT 13

ID W13905 standard; Protein; 1539 AA.

AC W13905;

DT 10-JUN-1997 (first entry)

DE Human SMCY protein.

KW SMCY; Y chromosome; male fertility; sex determination; H-Y antigen;

KW minor histocompatibility antigen; male contraceptive; evolution;

KW embryogenesis; spermatogenesis; tumour.

OS Homo sapiens.

PN W09710267-A1.

PD 20-MAR-1997.

PF 13-SEP-1996; U14547.

PR 14-SEP-1995; US-003744.

PR 07-MAR-1996; US-012973.

PA (PROM-) PROMEGA CORP.

PI Agulnik AI, Kent MG;

DR WPI: 97-202180/18.

DR N-PSDB; T60078.

PT New isolated male-specific human SMCY DNA - used to develop prods.

PT for use in e.g. embryonic, evolutionary, spermatogenesis,

PT tumorigenesis and histocompatibility studies

PS Claim 10; Page 30-37; 58pp; English.

CC The amino acid sequence (W13905) for SMCY, a protein widely

CC expressed in male tissues, was deduced from a composite cDNA

CC sequence (T60078) obt'd. from clones isolated from male lymphocyte

CC and testis cDNA libraries. Recombinant SMCY protein can be

CC produced in host cells and used to develop prods. for use in

CC embryonic, evolutionary, spermatogenesis, tumorigenesis and

CC histocompatibility studies, such as characterisation of the male-

CC specific H-Y antigen. Antibodies to the SMCY protein can be

CC used in immunological testing, and in a sex selection and screening

CC system for humans and non-human species.

CC Sequence 1539 AA;

Query Match

Best Local Similarity 4.3%; Score 92; DB 21; Length 1539;

Matches 11; Conservative 12; Mismatches 18; Indels 1; Gaps 1;

Db 304 knhsaqqidsyicqcsrgdednkllfdgdcddnyhifcll 345

| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 141 KINCSDGYIEYICRGAERVKRGLSFYSG-HSSFSSMYCML 181

RESULT 14

ID R34468 standard; Protein; 3011 AA.

AC R34468;

DT 30-JUL-1993 (first entry)

DE Encoded by full-length Hepatitis C virus clone JKI-B.

KW HCV; non-A, non-B hepatitis virus; NANBHV; liver disease;

KW Polymerase chain reaction; diagnostic method.

OS Hepatitis C virus.

FT Key Location/Qualifiers

FT Misc.difference 2414

FT /note= "not defined"

PN J05068562-A.

PD 23-MAR-1993.

PF 30-MAY-1991; 153736.

PR 30-MAY-1991; JP-153736.

PA (SANG) SANWA KAGAKU KENKYUSHO CO.

DR WPI: 93-130638/16.

DR N-PSDB; Q40426.

PT DNA and cDNA of hepatitis C virus - useful as probes for

PT diagnosing HCV infection

PS Claim 3; Page 6-18; 44pp; Japanese.

CC cDNA was prepared from HCV genomic RNA. Full-length clone JKI-B

CC and 14 shorter clones were isolated by PCR amplification.

CC Primer/probes derived from the sequences of these clones can be used

CC in diagnostic assays for HCV. See also Q40425-Q40439.

CC Sequence 3011 AA;

Query Match 4.3%; Score 91; DB 7; Length 3011;

Best Local Similarity 35.5%; Pred. No. 3.51e+01;

Matches 11; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

Db 364 mvgnwakvllmllfagvgdgttyvavghasq 394

| : : : : : | : : : : : | : : : : : | : : : : : |

QY 191 MKGDWARLLRPTLQGLVAVSIYVGLSRVSD 221

RESULT 15

ID W20704 standard; Protein; 170 AA.

AC W20704;

DT 15-JUL-1997 (first entry)

DE H. pylori transmembrane protein, O5apl1505orf10.

KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;

KW identification; binding compound; bacterium; life cycle; activator;

KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;

KW diagnosis.

OS Helicobacter pylori.

PN W09640893-A1.

PD 19-DEC-1996.

PF 06-JUN-1996; U09122.

PR 07-JUN-1995; US-487032.

PR 01-APR-1996; US-630405.

PA (ASTR) ASTRA AB.

PI Berglindh OT, Smith D, Mellgaard BL;

DR WPI: 97-022306/05.

DR N-PSDB; T67957.

PT Helicobacter pylori nucleic acid sequences and related

PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori

PT infection, and to detect Helicobacter

PS Claim 73; Page 1122; 1481pp; English.

CC This sequence represents a H. pylori protein likely to contain four

CC membrane spanning regions.

CC The protein may be used in a vaccine to prevent or treat H. pylori

CC infection or to identify H. pylori polypeptide binding compounds,

CC useful as potential H. pylori life cycle activators or inhibitors.

CC The genomic sequence of H. pylori (ATCC 55679) was determined from

CC overlapping contigs generated by mechanically shearing the bacterial

CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,

CC and the predicted coding regions defined by computer evaluation. To

CC identify likely H. pylori antigens for vaccine development, the amino

CC acid sequences predicted from various ORF were analysed for significant

CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from *H. pylori* by PCR amplification for recombinant polypeptide
CC production, e.g. in *E. coli* hosts.
SQ Sequence 170 AA;

Query Match 4.28; Score 88; DB 21; Length 170;
Best Local Similarity 31.8%; Pred. No. 5.62e+01;
Matches 21; Conservative 19; Mismatches 20; Indels 6; Gaps 6;
Db 41 qltsptak-gilmv-iflglaifiwknlrdwkeilmvylalkevpmlvaivsnl-kei 97
QY 197 RLRLPTLQFLGVAVSIIYGLSR-VSDYKHHSOVLTLGLOGALVAI-LVAVIVSDPFKER 254
Db 98 ss-kek 102
QY 255 TSPKER 260

Search completed: Tue Nov 4 10:24:46 1997
Job time : 74 secs.

This Page Blank (uspto)

W P S R L A
(TM)

Release 2.1D John F. Collins, BioComputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Nov 4 10:21:54 1997; MasPar time 10.95 Seconds
Tabular output not generated.
751.945 Million cell updates/sec

Title: >US-08-842-827-4
Description: (1-285) from US08842827.pap
Perfect Score: 2119
Sequence: 1 MFDRTRLPYVALDVLVLLA.....HTTLHETPTTGNHPSNHQP 285

Scoring table: PAM 150
Gap 11

Searched: 91006 seqs, 28888923 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir51
1:ann1 2:ann2 3:ann3 4:ann4 5:unann1 6:unann2 7:unann3
8:unann4 9:unann5 10:unann6 11:unann7 12:unann8
13:unann9 14:unann10 15:unenc 16:unrev

Statistics: Mean 46.052; Variance 108.101; scale 0.426

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length	ID	Description
1	193	9.1	274	11	S69561 hypothetical protein
2	181	8.5	289	11	S70114 hypothetical protein
3	121	5.7	345	1	QXKL2M NADH dehydrogenase (
4	117	5.5	216	10	S32217 hypothetical protein
5	111	5.2	608	10	S28036 penicillin-binding p
6	105	5.0	430	9	S15308 hypothetical protein
7	106	5.0	441	12	S28760 NADH dehydrogenase (
8	105	5.0	719	10	S28037 penicillin-binding p
9	105	5.0	719	10	A42893 penicillin-binding p
10	105	5.0	719	10	S28038 penicillin-binding p
11	104	4.9	552	11	J61173 cycloheximide resist
12	99	4.7	157	13	S58012 probable olfactory r
13	100	4.7	261	8	I38164 hypothetical protein
14	99	4.7	342	16	S68129 NADH dehydrogenase (
15	99	4.7	352	5	B34284 NADH dehydrogenase (
16	99	4.7	368	9	S54809 linoleoyl-CoA desatu
17	100	4.7	719	10	S28034 penicillin-binding p
18	100	4.7	719	10	S28031 penicillin-binding p
19	100	4.7	719	10	S28033 penicillin-binding p
20	100	4.7	719	10	S28032 penicillin-binding p
21	100	4.7	3388	4	GNWVDP genome polyprotein -
22					2.74e+00

22	98	4.6	271	9	I55244	adenosinetriphosphat	4.47e+00
23	97	4.6	346	5	S10188	NADH dehydrogenase (5.70e+00
24	97	4.6	366	10	B64413	hypothetical protein	5.70e+00
25	98	4.6	676	5	S54750	cycK protein - Rhizo	4.47e+00
26	97	4.6	723	10	D42707	H+/K+-transporting A	5.70e+00
27	97	4.6	885	8	A25817	nonstructural protei	5.70e+00
28	97	4.6	1127	11	S47445	MDM1 protein - yeast	5.70e+00
29	98	4.6	3391	4	GNWV16	genome polyprotein -	4.47e+00
30	98	4.6	3391	4	GNWV26	genome polyprotein -	4.47e+00
31	97	4.6	3391	8	JS0219	polyprotein - denque	5.70e+00
32	95	4.5	161	9	B64165	hypothetical protein	9.20e+00
33	95	4.5	346	5	S25423	NADH dehydrogenase (9.20e+00
34	96	4.5	347	5	S26152	NADH dehydrogenase (7.25e+00
35	96	4.5	347	5	S41836	NADH dehydrogenase (7.25e+00
36	95	4.5	366	7	I49519	bradykinin receptor	9.20e+00
37	95	4.5	366	16	I49525	bradykinin B2 recept	9.20e+00
38	96	4.5	397	16	S60334	purinoceptor P2X - r	7.25e+00
39	96	4.5	397	14	I58099	gene P2X3 protein -	7.25e+00
40	96	4.5	520	4	MXRW7	nonstructural protei	7.25e+00
41	96	4.5	615	14	S43285	noradrenaline transp	7.25e+00
42	95	4.5	727	9	A32561	probable cadmium-tra	9.20e+00
43	94	4.4	418	13	A46226	somatostatin recept	1.17e+01
44	94	4.4	936	8	I40711	sapB protein - Campy	1.17e+01
45	94	4.4	948	2	PXM0P2	H+-transporting ATPa	1.17e+01

ALIGNMENTS

RESULT	1	S69561	#type complete
ENTRY		hypothetical protein YDR503c - yeast (Saccharomyces cerevisiae)	
TITLE		#formal_name Saccharomyces cerevisiae	
ORGANISM		22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change	
DATE		06-Sep-1996	
ACCESSIONS		S69561	
REFERENCE		S69553	
#authors		Dietrich, F.S.	
#submission		submitted to the EMBL Data Library, August 1995	
#description		The sequence of S. cerevisiae cosmids 8166, 9787, 9717, and lambda 3073.	
#accession		S69561	
#molecule_type		DNA	
#residues		1-274	#label DIE
#cross-references		EMBL:U33057	
GENETICS			
#map_position		4R	
SUMMARY		#length 274 #molecular-weight 31586 #checksum 1196	

Query Match	9.1%;	Score 193;	DB 11;	Length 274;
Best Local Similarity	35.1%;	Pred. No. 2.63e-12;		
Matches	53;	Conservative 33;	Mismatches 56;	Indels 9; Gaps 9;
Db	107	isndf-hfhtsilclmlisinaaltgalkliignlrfpfdvrcidpdlqkmsdsdlvf	165	
Qy	91	ISNNYIATIKYIKAGTFLFGAAASQSLDIATAKYSIGRLRPHFLDVCDDPWSKINGSDGYI-	149	
Db	166	gldickqtnkwllyeglkstpsghsfivstmgtf-t-ylwgrv-ft-trntsriscwpll	221	
Qy	150	EYTCRGNARVK-EGRLSFYSYGHSSFSMYCMLFVALYLOARMKGDWARLLRPTLQFGLV	208	
Db	222	al-v-vmvsvrvidhrhwydvvgavlaflv	250	
Qy	209	AVSIYVGLSRVSDYKHHWSDVLTGLIOGALV	239	
RESULT	2	S70114	#type complete	
ENTRY		hypothetical protein YDR284c - yeast (Saccharomyces cerevisiae)		
TITLE		#formal_name Saccharomyces cerevisiae		
ORGANISM		24-Aug-1996 #sequence_revision 06-Sep-1996	#text_change	
DATE		06-Sep-1996		

```

ACCESSIONS S70114 #type complete
REFERENCE S70114 #type complete
#authors Fulton, L. #type complete
#submission submitted to the EMBL Data Library, May 1996
#description The sequence of S. cerevisiae cosmid 9819.
#accession S70114
#molecule_type DNA
#residues 1-289 #label FUL
#cross-references EMBL:U51031
GENETICS
#map_position 4R
SUMMARY #length 289 #molecular-weight 33514 #checksum 4703
Query Match 8.5%; Score 181; DB 11; Length 289;
Best Local Similarity 27.6%; Pred. No. 1.35e-10;
Matches 68; Conservative 62; Mismatches 98; Indels 18; Gaps 16;
Db 40 pferqfndltishyattervnmmlfv-yfsvps--lt-illigslladrhlifi 95
QY 34 PFORGFFCKDINSINYPHDSTAASVTLILVGVGLPVSSIIILGTLGVYCNLLHSNFSIN 93
Db 96 lytsllglsl-aw-f--sts-fftnfkwnwgrlprdfldrcqpv-eglpdltlftakdv 149
QY 94 NYIATYKAIGTFLFGAASQSTDTAKISIGRLRHFLDVCDDPSKINCSGYIEYII 153
Db 150 ctknherldgrtpt-sghssesfagilylyfwlcgqltlesplmrkwnvaflp11 208
QY 154 CRG-NAERVKEG-RLSFYSGHSSFSVMCMFLFVALYLQARMKGDWARL-L-RPTLQF-GLV 208
Db 209 gaal-talstqqrhfhvdfvlgslgymahffyrirfppidpdpfxplmd-dsdvt 266
QY 209 AVSIYVGLSRVSDYKHHWSVLTGLIQLGALVAILVAVYVSDFFKERTSFKERKESHTT 268
Db 267 leeaft 272
QY 269 LHETPT 274
RESULT 3
ENTRY QXXL2M #type complete
TITLE NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 -
ALTERNATE_NAMES African Cloned frog mitochondrion (SGC1)
ORGANISM NADH-ubiquinone oxidoreductase chain 2
#formal_name Mitochondrion Xenopus laevis #common_name
#accession African Cloned frog
DATE 28-Feb-1986 #sequence_revision 28-Feb-1986 #text_change
ACCESSIONS A00417
REFERENCE A00417
#authors Roe, B.A.; Ma, D.P.; Wilson, R.K.; Wong, J.F.H.
#journal J. Biol. Chem. (1985) 260:9759-9774
#title The complete nucleotide sequence of the Xenopus laevis
#cross-references MIM:85261388
#accession A00417
#molecule_type DNA
#residues 1-345 #label ROE
GENETICS
#genome mitochondrion
#genetic_code SGC1
CLASSIFICATION #superfamily NADH dehydrogenase (ubiquinone) chain 2
KEYWORDS membrane-associated complex; mitochondrion; NAD; oxidative
phosphorylation; oxidoreductase; respiratory chain
SUMMARY #length 345 #molecular-weight 37671 #checksum 3982
Query Match 5.7%; Score 121; DB 1; Length 345;
Best Local Similarity 25.3%; Pred. No. 1.11e-02;
Matches 21; Conservative 27; Mismatches 28; Indels 7; Gaps 7;
Db 51 raleastkylftgaasal-llfssln-nawltgwsildtnplscatmtaiaimklg- 107
QY 160 RVKEGRLSYSGHSSFSVMCMFLFVALYLQARMKGDWARL-LRPTLQFGLVAVSIYVGLSR 218

```

```

Db 108 lapf-hfwlpevlqslttgli 129
QY 219 VSDYKHHW-SDVLTGL-IOGALV 239
RESULT 4
ENTRY #type complete
TITLE hypothetical protein 2 - Bacillus megaterium
ORGANISM #formal_name Bacillus megaterium
DATE 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change
ACCESSIONS S32217
REFERENCE S32217
#authors Rauschenbach, R.; Isernhaagen, M.; Noeske-Jungblut, C.;
#submission Boldol, W.; Sievert, G.
#description submitted to the EMBL Data Library, March 1993
#accession Cloning, sequencing and expression of the gene for cytochrome
P450meg, the steroid 15beta-monooxygenase from Bacillus
megaterium ATCC 13368.
#accession S32217
#status preliminary
#molecule_type DNA
#residues 1-216 #label RAU
#cross-references EMBL:Z21972
SUMMARY #length 216 #molecular-weight 24946 #checksum 8213
Query Match 5.5%; Score 117; DB 10; Length 216;
Best Local Similarity 29.4%; Pred. No. 3.31e-02;
Matches 32; Conservative 31; Mismatches 37; Indels 9; Gaps 8;
Db 114 rpdhlridigysfsgpshammafsllygil-tfl-lwrhitawaril-lll-fgmlmi- 168
QY 155 RGNARVKE-GRLSFYSGHS--SFSMCMFLFVALYLQARMKGDWARLRLPTLQFGLVAVS 211
Db 169 lsigrlrlygvhypsdiagylagg-cwalsiwhffqyqdrknkdr 216
QY 212 IYVGLSRVSDYKHHWSVLTGLIQLGALVAILVAVYVSDFFKERTSFKER 260
RESULT 5
ENTRY S28036 #type fragment
TITLE penicillin-binding protein 1a - Streptococcus pneumoniae
ALTERNATE_NAMES (strain 8250) (fragment)
ORGANISM #formal_name Streptococcus pneumoniae
DATE 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
ACCESSIONS S28036
REFERENCE S28036
#authors Martin, C.; Sibbold, C.; Hakenbeck, R.
#journal EMBO J. (1992) 11:3831-3836
#title Relatedness of penicillin-binding protein 1a genes from
different clones of penicillin-resistant Streptococcus
pneumoniae isolated in South Africa and Spain.
#accession S28036
#status translation not shown
#molecule_type DNA
#residues 1-608 #label MAR
#cross-references EMBL:X57871
SUMMARY #length 608 #checksum 6599
Query Match 5.2%; Score 111; DB 10; Length 608;
Best Local Similarity 30.2%; Pred. No. 1.65e-01;
Matches 19; Conservative 19; Mismatches 22; Indels 3; Gaps 3;
Db 408 vynydrdyf-gnitlqyalqqrnpavetlqkvlrnraktfl-nglglidyplshysnai 465
QY 32 IYFQGFCKDINSINYPHDSTAASVTLILVGVGLPVSSIIILGTLGV-YNCLLHSNSF 90
Db 466 ssn 468
QY 91 ISN 93

```

```

RESULT 6
ENTRY S15308 #type complete
TITLE hypothetical protein 12.8 - Salmonella typhimurium
ORGANISM #formal_name Salmonella typhimurium
DATE 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
10-Nov-1995
ACCESSIONS S15308
REFERENCE S15296
#authors Jiang, X.M.; Neal, B.; Santiago, F.; Lee, S.J.; Romana, L.K.;
Reeves, P.R.
#journal Mol. Microbiol. (1991) 5:695-713
#title Structure and sequence of the rfb (O antigen) gene cluster of
Salmonella serovar typhimurium (strain LT2).
#cross-references M01D:91260454
#accession S15308
#status preliminary
#molecule_type DNA
#residues 1-430 #label MOL
#length 430 #molecular-weight 48627 #checksum 6054
SUMMARY
Query Match 5.0%; Score 105; DB 9; Length 430;
Best Local Similarity 31.1%; Pred. No. 7.82e-01;
Matches 19; Conservative 14; Mismatches 27; Indels 1; Gaps 1;
Db 178 yiyrgisvdklsilvlpvgmislcyyvryklyhvt-tkshyialrrssgffl 236
Qy 48 YPYHDSAASTVLIIVGVLPVSSIIIGETLSVCNLLHNSFSISNNYIATYKAIGTFL 107
Db 237 f 237
Qy 108 F 108

RESULT 7
ENTRY S28760 #type fragment
TITLE NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - blue
mussel mitochondrion (SGC4) (fragment)
ORGANISM #formal_name mitochondrion Mytilus edulis #common_name blue
mussel
DATE 20-Feb-1995 #sequence_revision 26-Jul-1996 #text_change
26-Jul-1996
ACCESSIONS S28760
REFERENCE S28743
#authors Hofmann, R.J.; Boore, J.L.; Brown, W.M.
#journal Genetics (1992) 131:397-412
#title A novel mitochondrial genome organization for the blue
mussel, Mytilus edulis.
#accession S28760
#molecule_type DNA
#residues 1-441 #label BRO
#cross-references EMBL:M83762
GENETICS
#gene ND5
#genome mitochondrion
#genetic_code SGC4
KEYWORDS membrane-associated complex; mitochondrion; NAD; oxidative
phosphorylation; oxidoreductase; respiratory chain
SUMMARY #length 441 #checksum 3300
Query Match 5.0%; Score 106; DB 12; Length 441;
Best Local Similarity 28.8%; Pred. No. 6.06e-01;
Matches 23; Conservative 25; Mismatches 28; Indels 4; Gaps 4;
Db 330 lpvsmgmsvaivsaga-pfmsgfskdlmlelmsdssvtygcyllelgl-lftsfy 387
Qy 15 LCVLLASMPMAVLKUGQYIPQRGFFCKDINSINYPYHDSAASTVLIIVGVLPVSSIIIL 74
Db 388 sari-vfsvmlgsn-yvnns 405
Qy 75 GETLSVCNLLHNSFSISNN 94
RESULT 8
ENTRY S28037 #type fragment
TITLE penicillin-binding protein 1a - Streptococcus pneumoniae
ORGANISM #formal_name Streptococcus pneumoniae
DATE 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
18-Nov-1994
ACCESSIONS S28037
REFERENCE S28031
#authors Martin, C.; Sibold, C.; Hakenbeck, R.
#journal EMBO J. (1992) 11:3831-3836
#title Relatedness of penicillin-binding protein 1a genes from
different clones of penicillin-resistant Streptococcus
pneumoniae isolated in South Africa and Spain.
#accession S28037
#status translation not shown
#molecule_type DNA
#residues 1-719 #label MAR
#cross-references EMBL:X67872
SUMMARY #length 719 #checksum 2230
Query Match 5.0%; Score 105; DB 10; Length 719;
Best Local Similarity 28.6%; Pred. No. 7.82e-01;
Matches 18; Conservative 20; Mismatches 22; Indels 3; Gaps 3;
Db 408 vynydrgyf-gnitlqvalqsrnrvpavetlnkvlnrktfl-nglglidyphsina 465
Qy 32 IYPTQRGFFCKDINSINYPYHDSAASTVLIIVGVLPVSSIIIGETLSV-YCNLLHNSF 90
Db 466 ssn 468
Qy 91 ISN 93

RESULT 9
ENTRY A42893 #type complete
TITLE penicillin-binding protein 1A - Streptococcus pneumoniae
ORGANISM #formal_name Streptococcus pneumoniae
DATE 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change
07-Apr-1994
ACCESSIONS A42893
REFERENCE A42893
#authors Martin, C.; Briese, T.; Hakenbeck, R.
#journal J. Bacteriol. (1992) 174:4517-4523
#title Nucleotide sequences of genes encoding penicillin-binding
proteins from Streptococcus pneumoniae and Streptococcus
oralis with high homology to Escherichia coli
penicillin-binding proteins 1A and 1B.
#accession A42893
#status preliminary
#molecule_type DNA
#residues 1-719 #label MAR
#cross-references GB:M90527
SUMMARY #length 719 #molecular-weight 79701 #checksum 2518
Query Match 5.0%; Score 105; DB 10; Length 719;
Best Local Similarity 28.6%; Pred. No. 7.82e-01;
Matches 18; Conservative 20; Mismatches 22; Indels 3; Gaps 3;
Db 408 vynydrgyf-gnitlqvalqsrnrvpavetlnkvlnrktfl-nglglidyphsina 465
Qy 32 IYPTQRGFFCKDINSINYPYHDSAASTVLIIVGVLPVSSIIIGETLSV-YCNLLHNSF 90
Db 466 ssn 468
Qy 91 ISN 93

RESULT 10
ENTRY S28038 #type fragment
TITLE penicillin-binding protein 1a - Streptococcus pneumoniae
ORGANISM #formal_name Streptococcus pneumoniae
DATE 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
17-Apr-1993

```


#title The complete nucleotide sequence of the mitochondrial genome of the lungfish (Protopterus dolloi) supports its phylogenetic position as a close relative of land vertebrates.

#accession S68129
#status preliminary
#residues 1-342 #label ZAR
#cross-references EMBL:L42813

SUMMARY #length 342 #molecular-weight 37535 #checksum 390

Query Match 4.7%; Score 99; DB 16; Length 342;
Best Local Similarity 25.0%; Pred. No. 3.50e+00;
Matches 22; Conservative 32; Mismatches 25; Indels 9; Gaps 8;

Db 51 rateaatkyflaqaasim-ilfssm-inawvagewnitnlsptsa-tliltlalaikig 107

QY 160 RVKEGRLSFYSGHSSFSMYCMLFVALYLQARMKGDW--ARLLRPTLQFLGLVAVSIYVGLS 217

Db 108 -lap-mhfwlpevlqgvtlmt-gailvt 132

QY 218 RVSDYKHHW-SDVLTGLTQALVA 244

RESULT 15

ENTRY B34284 #type complete
TITLE NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 - sea urchin (Paracentrotus lividus) mitochondrion (SGC8)
ORGANISM #formal_name mitochondrion Paracentrotus lividus #common_name common urchin

DATE 29-Jun-1990 #sequence_revision 29-Jun-1990 #text_change 04-Nov-1994

ACCESSIONS B34284; A26510

REFERENCE A34284

#authors Cantatore, P.; Roberti, M.; Rainaldi, G.; Gadaleta, M.N.; Sacccone, C.

#journal J. Biol. Chem. (1989) 264:10965-10975

#title The complete nucleotide sequence, gene organization, and genetic code of the mitochondrial genome of Paracentrotus lividus.

#cross-references MUID:89291831

#accession B34284

#molecule_type DNA

#residues 1-352 #label CAN

#cross-references GB:J04815

REFERENCE A26510

#authors Cantatore, P.; Roberti, M.; Morisco, P.; Rainaldi, G.; Gadaleta, M.N.; Sacccone, C.

#journal Gene (1987) 53:41-54

#title A novel gene order in the Paracentrotus lividus mitochondrial genome.

#cross-references MUID:87248108

#accession A26510

#molecule_type DNA

#residues 134-135, 'M', 137-148, 'K', 150-154, 'K', 156-183, 'M', 185-189, 'M', 191-193, 202-211, 'K', 213-223, 'M', 225-230, 'K', 232-238, 'M', 240, 'M', 242-247, 'M', 249-263 #label CA2

#cross-references GB:M16516
#note the authors translated the nucleotide sequence using the standard genetic code; the authors translation is shown

GENETICS

#gene ND2

#genome mitochondrion

#genetic_code SGC8

CLASSIFICATION #superfamily NADH dehydrogenase (ubiquinone) chain 2

KEYWORDS membrane-associated complex; mitochondrion; NAD; oxidoreductase

SUMMARY #length 352 #molecular-weight 39009 #checksum 8650

Query Match

Best Local Similarity 4.7%; Score 99; DB 5; Length 352;

Matches 28; Conservative 22; Mismatches 26; Indels 10; Gaps 9;

Db 51 rnleatikyflvqa-fsaalllnsal-igawfgsgswali-pmesfpclclslvalafnlg 107

QY 160 RVKEGRLSFYSGHSSFSMYCMLFVALYLQARMKGDWARLLRPTLQF-GL-VAVSIYVGLS 217

Db 108 -laac-hfwlpdvlggipfiggliia 131

QY 218 RVSDYKHHW-SDVLTGL--IQGALVA 240

Search completed: Tue Nov 4 10:23:14 1997

Job time : 80 secs.

This Page Blank (uspto)

W E S R E L H
(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by Intelligenetics, Inc.

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Nov 4 10:20:41 1997; MasPar time 7.75 seconds
780.375 Million cell updates/sec
Tabular output not generated.

Title: >US-08-842-827-4
Description: (1-285) from US08842827.ppe
Perfect Score: 2119
Sequence: 1 MFDRTRLPYVALDVLCVLLA.....HTTLHETPTTGNHPSNHQP 285

Scoring table: PAM 150
Gap 11

Searched: 59021 seqs, 21210388 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot34
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11

Statistics: Mean 47.899; Variance 90.096; scale 0.532

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	454	21.4	341	11	YSX3_CAEEL	HYPOTHETICAL 39.0 KD
2	121	5.7	345	7	NU2M_YENLA	NADH-UBIQUINONE OXIDO
3	105	5.0	430	8	RFBX_SALTY	RFBX PROTEIN.
4	106	5.0	522	7	NU5M_MYTED	NADH-UBIQUINONE OXIDO
5	105	5.0	719	7	PBPA_STPRN	PENICILLIN-BINDING PR
6	104	4.9	552	3	CYHR_CANMA	CYCLOHEXIMIDE RESISTA
7	101	4.8	345	11	YD99_SCHPO	PUTATIVE MITOCHONDRIA
8	100	4.7	261	11	YCB9_PSEDE	HYPOTHETICAL 27.4 KD
9	99	4.7	352	7	NU2M_PARLI	NADH-UBIQUINONE OXIDO
10	100	4.7	3388	7	POLG_DEN2P	GENOME POLYPROTEIN (C
11	97	4.6	346	6	NU2M_CHICK	NADH-UBIQUINONE OXIDO
12	97	4.6	443	6	MDM1_YEAST	STRUCTURAL PROTEIN MD
13	98	4.6	676	2	CCWF_RHIME	CYTOCHROME C-TYPE BIO
14	98	4.6	704	7	PM22_LFCEB	PLASMA MEMBRANE ATPAS
15	97	4.6	723	2	CADA_BACFI	PROBABLE CADMIUM-TRAN
16	98	4.6	3391	7	POLG_DEN26	GENOME POLYPROTEIN (C
17	98	4.6	3391	7	POLG_DEN27	GENOME POLYPROTEIN (C
18	97	4.6	3391	7	POLG_DEN2N	GENOME POLYPROTEIN (C
19	95	4.5	161	11	YIAM_HAEIN	HYPOTHETICAL PROTEIN
20	95	4.5	346	6	NU2M_COTJA	NADH-UBIQUINONE OXIDO
21	96	4.5	347	7	NU2M_PROVI	NADH-UBIQUINONE OXIDO
22	96	4.5	347	7	NU2M_HALGR	NADH-UBIQUINONE OXIDO

23	95	4.5	366	2	BBR2_MOUSE	B2 BRADYKININ RECEPTOR	2.20e+00
24	96	4.5	397	7	P2X3_RAT	P2X PURINOCEPTOR 3 (A	1.66e+00
25	96	4.5	520	10	VP7_WTV	NONSTRUCTURAL PROTEIN	1.66e+00
26	96	4.5	615	6	NANO_BOVIN	SODIUM-DEPENDENT NORA	1.66e+00
27	95	4.5	727	2	CADA_STAAU	PROBABLE CADMIUM-TRAN	2.20e+00
28	95	4.5	738	10	VU47_HSV6Z	GLYCOPROTEIN U47 PREC	2.20e+00
29	93	4.4	274	1	BACH_HALSP	HALORHODOPSIN PRECURS	3.58e+00
30	94	4.4	418	9	SSR3_HUMAN	SOMATOSTATIN RECEPTOR	2.91e+00
31	93	4.4	533	11	YADC_SCHPO	HYPOTHETICAL 62.2 KD	3.85e+00
32	94	4.4	804	2	CADD_STAAU	PROBABLE CADMIUM-TRAN	2.91e+00
33	94	4.4	947	7	PN42_ARATH	PLASMA MEMBRANE ATPAS	2.91e+00
34	94	4.4	1073	5	ITA6_HUMAN	INTEGRIN ALPHA-6 PREC	2.91e+00
35	92	4.3	113	5	IMMO_ECOLI	IMMUNITY PROTEIN FOR	5.08e+00
36	91	4.3	271	1	ATP6_ECOLI	ATP SYNTHASE A CHAIN	6.88e+00
37	91	4.3	387	10	V748_NPVAC	P48 PROTEIN (P45)	6.68e+00
38	91	4.3	590	2	CHLL_ARATH	NITRATE/CHLORATE TRAN	6.68e+00
39	92	4.3	637	7	PBPA_STROK	PENICILLIN-BINDING PR	5.08e+00
40	91	4.3	652	7	NU5M_PODAN	NADH-UBIQUINONE OXIDO	6.68e+00
41	90	4.2	82	11	YCN2_LACLA	HYPOTHETICAL 9.7 KD P	8.76e+00
42	90	4.2	532	11	YH7_YEAST	HYPOTHETICAL 61.2 KD	8.76e+00
43	90	4.2	561	10	VGIF_SV41	FUSION GLYCOPROTEIN P	8.76e+00
44	90	4.2	637	7	NU5M_STRPU	NADH-UBIQUINONE OXIDO	8.76e+00
45	90	4.2	1463	7	PA2R_BOVIN	180 KD SECRETORY PHOS	8.76e+00

ALIGNMENTS

RESULT	1
ID	YSX3_CAEEL
AC	Q10022; STANDARD; PRT; 341 AA.
DT	01-OCT-1996 (REL. 34, CREATED)
DT	01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT	01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE	HYPOTHETICAL 39.0 KD PROTEIN T28D9.3 IN CHROMOSOME II.
GN	T28D9.3.
OS	CAENORHABDITIS ELEGANS.
OC	EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	SPRAIN-BRISTOL N2;
RA	WATERSTON R.;
RL	SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC	-1- SIMILARITY: WEAK, TO YEAST D9719.9.
DR	EMBL: U28738; G861267; -
DR	WORMPEP; T28D9.3; CE02068.
KW	HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT	TRANSMEM 30 50 POTENTIAL.
FT	TRANSMEM 71 91 POTENTIAL.
FT	TRANSMEM 122 142 POTENTIAL.
FT	TRANSMEM 223 243 POTENTIAL.
FT	TRANSMEM 257 277 POTENTIAL.
SQ	SEQUENCE 341 AA; 39028 MW; 89AE6E81 CRC32;

Query Match 21.4%; Score 454; DB 11; Length 341;
Best Local Similarity 32.8%; Pred. No. 7.00e-67;
Matches 82; Conservative 66; Mismatches 90; Indels 12; Gaps 11;

Db	34	flataavtvtvptllgvsqrgffoddsiryeayrkdttitavqlmlynlvnaatvlfvey	93
QY	18	LLASMPMAVLKLGQIYFPQGFCKDINSINTPYHDSSTAASTVLILVGVGLPVSSIILGE-	76
Db	94	ymdqvesni-nnpyrwrnllhvlfrllityfgysigfmmialivtkhvgirlp	152
QY	77	-TL-SVYCNLLHSNFI-SNNYIATYI-KAIGTFLFGAA-SQSLT-DI-AKYSIGRLRP	129
Db	153	hflgvcklandtcvgdshrvtctgtgppelvarksfygshsavlcatwsalyi	212
QY	130	HFLDVCDDPWSK-INC-SDGYEYICRGNARVKEGRLSFYSGHSSFSMCMLEVALYL	187
Db	213	qarlgpvlrnnrivvpsqtlmfmiglgisfridtnkhwsdvlvgifigflavtctf	272
QY	188	QARMKGDNA-RLLRPTLQGLVAVSVIYVGLSRVSDYKHHWSVDLTGLIQGALVALVAVY	246


```

RN  [2]
RP  SEQUENCE FROM N.A.
RA  MEDLINE; 92325042.
RX  MARTIN C., BRIESE T., HAKENBECK R.;
RL  J. BACTERIOL. 174:4517-4523(1992).
RN  [3]
RP  SEQUENCE OF 293-369 FROM N.A.
RA  STRAIN-R6X;
RX  MEDLINE; 95020625.
RL  PEARCE B.J., YIN Y.B., MASURE H.R.;
RM  MOL. MICROBIOL. 9:1037-1050(1993).
CC  -1- FUNCTION: CELL WALL FORMATION.
CC  -1- THESE STRAINS ARE PENICILLIN-SENSITIVE.
CC  -1- PATHWAY: FINAL STAGES IN PEPTIDOGLYCAN SYNTHESIS.
CC  -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
DR  EMBL; X67873; G47420; -.
DR  EMBL; X67872; G47418; -.
DR  EMBL; M90527; G153768; -.
DR  PIR; S28038; S28038.
KW  PEPTIDOGLYCAN SYNTHESIS; ANTIBIOTIC RESISTANCE; CELL WALL;
KW  MULTIFUNCTIONAL ENZYME.
FT  ACT_SITE 370 370 ACYLATED BY PENICILLIN (BY SIMILARITY).
FT  DOMAIN 657 683 SER-RICH.
FT  VARIANT 124 124 T -> A (IN STRAIN R6).
FT  VARIANT 386 386 V -> I (IN STRAIN 63915).
FT  VARIANT 388 388 D -> E (IN STRAIN R6).
FT  VARIANT 397 397 E -> K (IN STRAIN 63915).
FT  VARIANT 523 523 M -> I (IN STRAIN 63915).
FT  VARIANT 533 533 E -> D (IN STRAINS 63915 AND R6).
FT  VARIANT 540 540 T -> S (IN STRAINS 63915 AND R6).
FT  VARIANT 657 657 S -> N (IN STRAIN 63915).
SQ  SEQUENCE 719 AA; 79745 MW; 627E3A7B CRC32;

Query Match 5.0%; Score 105; DB 7; Length 719;
Best Local Similarity 28.6%; Pred. No. 1.16e-01;
Matches 18; Conservative 20; Mismatches 22; Indels 3; Gaps 3;

Db 408 vynwrgyf-gnitqlvalqgnrnpavetlnkylgnraktfl-ngldvpslhytsnai 465
QY 32 IYFPGFCKNSINYPYHDSTAASVLLVGVGLPVSSIIIGETLSV-YCNLLHNSF 90
Db 466 ssn 468
QY 91 ISN 93

RESULT 6
ID CYHR_CANNA STANDARD; PRT; 552 AA.
AC P32071;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE CYCLOHEXIMIDE RESISTANCE PROTEIN.
GN CYHR.
OS CANDIDA MALTOSA (YEAST).
OC EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-VSB-889;
RM MEDLINE; 92331941.
RA SASNAUSKAS K., JOMANTIENE R., LEBEDIENE E., LEBEDYS J.,
RA JANUSKA A., JANULAITIS A.;
RL GENE 116:105-108(1992).
CC -1- FUNCTION: PROBABLE TRANSPORTER. CONFERS RESISTANCE TO
CC CYCLOHEXIMIDE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC AS THE DRUG RESISTANCE TRANSLOCASE FAMILY). BELONGS TO THE
CC CARL/CYHR SUBFAMILY.
DR EMBL; M64932; G170849; -.
DR PIR; JC1173; JC1173.
KW ANTIBIOTIC RESISTANCE; TRANSPORT; TRANSMEMBRANE.
FT TRANSMEM 100 120 POTENTIAL.

```

```

FT TRANSMEM 137 157 POTENTIAL.
FT TRANSMEM 168 188 POTENTIAL.
FT TRANSMEM 194 213 POTENTIAL.
FT TRANSMEM 225 246 POTENTIAL.
FT TRANSMEM 262 282 POTENTIAL.
FT TRANSMEM 346 362 POTENTIAL.
FT TRANSMEM 381 399 POTENTIAL.
FT TRANSMEM 419 439 POTENTIAL.
FT TRANSMEM 445 464 POTENTIAL.
FT TRANSMEM 477 494 POTENTIAL.
FT TRANSMEM 518 539 POTENTIAL.
SQ SEQUENCE 552 AA; 61366 MW; AEF45532 CRC32;

Query Match 4.9%; Score 104; DB 3; Length 552;
Best Local Similarity 25.7%; Pred. No. 1.57e-01;
Matches 18; Conservative 29; Mismatches 17; Indels 6; Gaps 5;

Db 450 ligmalfavgaflifqlfnymavsfkveylasvafsvsagafplfgralynnl 509
QY 62 LVGVGL-PVSSIIIGETLSVYCNLLHNSFISNNYIA-TYKAI--GTF-LFGAASQSL 116
Db 510 s-ldkfpvgw 518
QY 117 TDIKYSIGR 126

RESULT 7
ID YDE9_SCHPO STANDARD; PRT; 345 AA.
AC Q10442;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PUTATIVE MITOCHONDRIAL CARRIER PROTEIN C12B10.09.
GN SPAC12B10.09.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA BADCOCK K., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE (POTENTIAL).
CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.
DR EMBL; Z70721; E235727; -.
KW HYPOTHETICAL PROTEIN; MITOCHONDRION; INNER MEMBRANE; REPEAT;
KW TRANSMEMBRANE; TRANSPORT.
KW TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 83 103 POTENTIAL.
FT TRANSMEM 128 148 POTENTIAL.
FT TRANSMEM 220 240 POTENTIAL.
FT TRANSMEM 262 282 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 319 339 POTENTIAL.
SQ SEQUENCE 345 AA; 38451 MW; D059679A CRC32;

Query Match 4.8%; Score 101; DB 11; Length 345;
Best Local Similarity 28.7%; Pred. No. 3.86e-01;
Matches 29; Conservative 28; Mismatches 38; Indels 6; Gaps 5;

Db 93 glaydlslfpdltlkt--rlqakgfvknggfghvvglsilvgsagpsalf-fttyen 149
QY 66 GLPVS-SIIIGETLSVYCNLLHNSFISNNYIATYKAI--GTF-LFGAASQSLDIKY-S 123
Db 150 mksrlsqsglqldpqlhmcasasgeiaacivrvptevikq 190
QY 124 I-GRLRPHFLDVCDPDWSKINCSGDYIEYICRGNAERVKE 163

RESULT 8
ID YCB9_PSEDE STANDARD; PRT; 261 AA.
AC P29942;

```

```

DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DE DE HYPOPHETICAL 27.4 K D PROTEIN IN COBO 3'REGION (ORF9).
DE PSEUDOMONAS DENITRIFICANS.
OS PROKARYOTA: GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
OC PSEUDOMONADACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92011366.
RA CROUZET J., LEVY-SCHIL S., CAMERON B., CAUCHOIS L., RIGAUD S.,
RA ROUYEZ M.-C., BLANCHE F., DEBUSSCHE L., THIBAUT D.;
RL J. BACTERIOLOG. 173:6074-6087(1991).
CC -1- SUBCELLULAR LOCATION: MEMBRANE (PROBABLE).
DR EMBL; M62866; G551929; -.
DR PIR; I38164; I38164.
KW HYPOPHETICAL PROTEIN; MEMBRANE.
SQ SEQUENCE 261 AA; 27374 MW; 5D508436 CRC32;

Query Match 4.7%; Score 100; DB 11; Length 261;
Best Local Similarity 31.1%; Pred. No. 5.19e-01;
Matches 19; Conservative 18; Mismatches 21; Indels 3; Gaps 3;

Db 102 vlkaipfliaiaalyfglkpnmgdvqdh-srvtpfvttltlvp-ligfydvgfpgtgs 159
QY 198 LLRFTLQFLGVAVSIYVGL-SRVS DYKHHSVDLTGLTGLQALVAILVAVVVSDFKERTS 256
Db 160 f 160
QY 257 F 257

RESULT 9 STANDARD; PRT; 352 AA.
ID NU2M_PARLI
AC P12771;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DE 01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 2 (EC 1.6.5.3).
GN ND2.
OS PARACENTROTUS LIVIDUS (COMMON SEA URCHIN).
OC MITOCHONDRION.
OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
OC EUECHINOIDEA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89291831.
RA CANTATORE P., ROBERTI M., RAINALDI G., GADALETA M.N., SACCONI C.;
RL J. BIOL. CHEM. 264:10965-10975(1989).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
DR EMBL; J04815; G453499; -.
DR PIR; B34284; B34284.
KW OXIDOREDUCTASE; NAD; UBIQUINONE; MITOCHONDRION.
SQ SEQUENCE 352 AA; 39009 MW; 39BB4327 CRC32;

Query Match 4.7%; Score 99; DB 7; Length 352;
Best Local Similarity 32.6%; Pred. No. 6.96e-01;
Matches 28; Conservative 22; Mismatches 26; Indels 10; Gaps 9;

Db 51 rneatikylvqa-fsaallinsal-lqawfsgswsali-pmesfpalcisvalafnlg 107
QY 160 RVKEGRLSFYSCHSFSMYCMLFVALYLQARKMGDWALLRPTLQF-GL-VAVSIYVGLS 217
Db 108 -laac-hfwlpdvlpqglpfiqllia 131
QY 218 RVSDYKHWH-SDVLTGL--IQGALVA 240

RESULT 10 STANDARD; PRT; 3388 AA.
ID POLG_DEN2P
AC P12823;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DE 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN (CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
DE PROTEIN (ENVELOPE GLYCOPROTEIN M); MAJOR ENVELOPE PROTEIN E;
DE NONSTRUCTURAL PROTEINS NS1, NS2A, NS2B, NS4A AND NS4B; HELICASE (NS3);
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (NS5)).
OS DENGUE VIRUS TYPE 2 (STRAIN PR159/S1).
OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; FLAVIVIRIDAE;
OC FLAVIVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88101365.
RA HAHN Y.S., GALLER R., HUNKAPILLER T., DALRYMPLE J.M., STRAUSS J.H.,
RA STRAUSS E.G.;
RL VIROLOGY 162:167-180(1988).
RN [2]
RP C-TERMINAL OF M, E, AND NS1.
RC STRAIN-NEW-GUINEA;
RX MEDLINE; 89299482.
RA WRIGHT P.J., CAUCHI M.R., NG M.L.;
RL VIROLOGY 171:61-67(1989).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
DR EMBL; X19197; G323655; -.
DR PIR; A29972; GNWVDP.
KW POLYPROTEIN; GLYCOPROTEIN; RNA-DIRECTED RNA POLYMERASE; CORE PROTEIN;
KW COAT PROTEIN; ENVELOPE PROTEIN; HELICASE; ATP-BINDING; TRANSMEMBRANE;
KW NONSTRUCTURAL PROTEIN.
RN [1]
RP CAPSID PROTEIN C.
FT CHAIN 2 114
FT PROPEP 115 205
FT CHAIN 206 280
FT CHAIN 281 775
FT CHAIN 776 1127
FT CHAIN 1128 1345
FT CHAIN 1346 1475
FT CHAIN 1476 2090
FT CHAIN 2091 2376
FT CHAIN 2377 2488
FT CHAIN 2489 3388
FT NP_BIND 1668 1675
FT SITE 1759 1762
FT TRANSMEM 43 59
FT TRANSMEM 101 117
FT TRANSMEM 268 284
FT TRANSMEM 727 743
FT TRANSMEM 757 773
FT TRANSMEM 283 310
FT DISULFID 340 396
FT DISULFID 354 385
FT DISULFID 372 401
FT DISULFID 465 565
FT DISULFID 582 613
FT CARBOHYD 183 183
FT CARBOHYD 347 347
FT CARBOHYD 433 433
FT CARBOHYD 905 905
FT CARBOHYD 982 982
FT CARBOHYD 1134 1134
FT CARBOHYD 1174 1174
FT CARBOHYD 2298 2298
FT CARBOHYD 2302 2302
FT CARBOHYD 2454 2454
FT CARBOHYD 2482 2482
SQ SEQUENCE 3388 AA; 379213 MW; 8D0A2D62 CRC32;

Query Match 4.7%; Score 100; DB 7; Length 3388;
Best Local Similarity 30.2%; Pred. No. 5.19e-01;
Matches 19; Conservative 20; Mismatches 18; Indels 6; Gaps 6;

```

```

Db 1306 p1f1tsqqkadvipia-ltik-glntptaifltltsrtsk-krswp-lneainavgmvs1 1361
QY 184 ALYLQA-RMKGDWARLLRPTLQFLGVAIVSYG-LSRVSDYKHHWSDLVTGLTQIGALVAI 241

Db 1362 las 1364
QY 242 LVA 244

RESULT 11
ID NU2M-CHICK STANDARD; PRT; 346 AA.
AC P18937;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 2 (EC 1.6.5.3).
GN ND2.
OS GALLUS GALLUS (CHICKEN).
OG MITOCHONDRION.
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC GALLIFORMES.
RN [1]
RX MEDLINE: 90230301.
RA DESJARDINS P., MORAIS R.;
RL J. MOL. BIOL. 212:599-634 (1990).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.
DR EMBL: X52392; G12962; -.
DR PIR: S10188; G10188.
KW OXIDOREDUCTASE; NAD; UBIQUINONE; MITOCHONDRION.
SQ SEQUENCE 346 AA; 38333 MW; 60A7C67B CRC32;

Query Match 4.6%; Score 97; DB 6; Length 346;
Best Local Similarity 26.0%; Pred. No. 1.24e+00;
Matches 25; Conservative 33; Mismatches 28; Indels 10; Gaps 8;

Db 44 iskshpraateatikiyftgstasal-ilfsm-tnawstgwditqlnhtpscl-mltm 100
QY 153 ICGNAERVKRSLFSYSGHSFMYCMLFVALYLQARMKGDW--ARLLRPTLQFLGIVAV 210

Db 101 aiaiklg1vp-f-hfwfpevlqgssltallstlm 134
QY 211 SIYVGLSRVSDYKHHW-SDVLTG--LIQALVALLV 243

RESULT 12
ID MDM1 YEAST STANDARD; PRT; 443 AA.
AC Q01846;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE STRUCTURAL PROTEIN MDM1.
GN MDM1 OR YML104C OR YMR339, 15C.
OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE: 92332595.
RA MCCONNELL S.J., YAFFE M.P.;
RL J. CELL BIOL. 118:385-395(1992).
RN [2]

Query Match 4.6%; Score 97; DB 6; Length 346;
Best Local Similarity 26.0%; Pred. No. 1.24e+00;
Matches 25; Conservative 33; Mismatches 28; Indels 10; Gaps 8;

Db 101 aiaiklg1vp-f-hfwfpevlqgssltallstlm 134
QY 211 SIYVGLSRVSDYKHHW-SDVLTG--LIQALVALLV 243

RESULT 13
ID CCMP_RHIME STANDARD; PRT; 676 AA.
AC P45404;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE CYTOCHROME C-TYPE BIOGENESIS PROTEIN CYC.
GN CYC.
OS RHIZOBIUM MELILOTI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
OC RHIZOBIACEAE.
RN [1]
RX SEQUENCE FROM N.A.
RX STRAIN-AK631;
RA KERESZT A., SLASKA-KISS K., PUTNOKY P., BANFALVI Z., KONDOROSI A.;
RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: REQUIRED FOR THE BIOGENESIS OF C-TYPE CYTOCHROMES.
CC POSSIBLE SUBUNIT OF A HEME LYASE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
(POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CCMP/CYCK/CCLI/NRFE/YCF5 FAMILY.
KW CYTOCHROME C-TYPE BIOGENESIS; TRANSMEMBRANE; INNER MEMBRANE.
FT TRANSMEM 8 28
FT TRANSMEM 42 62
FT TRANSMEM 94 114
FT TRANSMEM 123 143
FT TRANSMEM 175 195
FT TRANSMEM 210 230
FT TRANSMEM 233 253
FT TRANSMEM 273 293
FT TRANSMEM 311 331
FT TRANSMEM 356 376
FT TRANSMEM 393 413
FT TRANSMEM 418 438
FT TRANSMEM 445 465
FT TRANSMEM 493 513
FT TRANSMEM 617 637
SQ SEQUENCE 676 AA; 72643 MW; C9947521 CRC32;

Query Match 4.6%; Score 98; DB 2; Length 676;
Best Local Similarity 35.3%; Pred. No. 9.31e-01;
Matches 18; Conservative 14; Mismatches 15; Indels 4; Gaps 4;

Db 176 lylgyvgfsv-cfsfavaalliegriadaavrvpwtlaawtftlaqiang 225
QY 168 FYSGHSFMYCMLF-VALYLQARMKGDWARLRP-TLQ-FGLVAVSIYVG 215

RESULT 14
ID PMA2 LYCES STANDARD; PRT; 704 AA.
AC P23980;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE PLASMA MEMBRANE ATPASE 2 (EC 3.6.1.35) (PROTON PUMP) (FRAGMENT).

```

```

DR SGD; L0001051; MDM1.
FT CONFLICT 88 88 Q -> R (IN REF. 1).
FT CONFLICT 215 215 T -> I (IN REF. 1).
FT CONFLICT 379 379 R -> K (IN REF. 1).
FT CONFLICT 397 397 T -> S (IN REF. 1).
FT CONFLICT 412 412 G -> R (IN REF. 1).
SQ SEQUENCE 443 AA; 51470 MW; 621A08C9 CRC32;

Query Match 4.6%; Score 97; DB 6; Length 443;
Best Local Similarity 34.4%; Pred. No. 1.24e+00;
Matches 11; Conservative 9; Mismatches 11; Indels 1; Gaps 1;

Db 280 pfysqpkfsvkqicldfslfalnkanagwlr 311
QY 167 SFYSGHSFMYCMLFVALYLQARMKGDWAR 197

RESULT 13
ID CCMP_RHIME STANDARD; PRT; 676 AA.
AC P45404;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE CYTOCHROME C-TYPE BIOGENESIS PROTEIN CYC.
GN CYC.
OS RHIZOBIUM MELILOTI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
OC RHIZOBIACEAE.
RN [1]
RX SEQUENCE FROM N.A.
RX STRAIN-AK631;
RA KERESZT A., SLASKA-KISS K., PUTNOKY P., BANFALVI Z., KONDOROSI A.;
RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: REQUIRED FOR THE BIOGENESIS OF C-TYPE CYTOCHROMES.
CC POSSIBLE SUBUNIT OF A HEME LYASE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
(POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CCMP/CYCK/CCLI/NRFE/YCF5 FAMILY.
KW CYTOCHROME C-TYPE BIOGENESIS; TRANSMEMBRANE; INNER MEMBRANE.
FT TRANSMEM 8 28
FT TRANSMEM 42 62
FT TRANSMEM 94 114
FT TRANSMEM 123 143
FT TRANSMEM 175 195
FT TRANSMEM 210 230
FT TRANSMEM 233 253
FT TRANSMEM 273 293
FT TRANSMEM 311 331
FT TRANSMEM 356 376
FT TRANSMEM 393 413
FT TRANSMEM 418 438
FT TRANSMEM 445 465
FT TRANSMEM 493 513
FT TRANSMEM 617 637
SQ SEQUENCE 676 AA; 72643 MW; C9947521 CRC32;

Query Match 4.6%; Score 98; DB 2; Length 676;
Best Local Similarity 35.3%; Pred. No. 9.31e-01;
Matches 18; Conservative 14; Mismatches 15; Indels 4; Gaps 4;

Db 176 lylgyvgfsv-cfsfavaalliegriadaavrvpwtlaawtftlaqiang 225
QY 168 FYSGHSFMYCMLF-VALYLQARMKGDWARLRP-TLQ-FGLVAVSIYVG 215

RESULT 14
ID PMA2 LYCES STANDARD; PRT; 704 AA.
AC P23980;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE PLASMA MEMBRANE ATPASE 2 (EC 3.6.1.35) (PROTON PUMP) (FRAGMENT).

```

```

GN LHA2.
OS LYCOPERSICON ESCULENTUM (TOMATO).
OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
OC SOLANALES; SOLANACEAE.
[1]
RN SEQUENCE FROM N.A.
RC TISSUE=ROOT;
RA EWING N.N., WIMMERS L.E., MEYER D.J., CHETELAT R.T., BENNETT A.B.;
RL PLANT PHYSIOL. 94:1874-1881(1990).
CC -!- FUNCTION: THE PLASMA MEMBRANE ATPASE OF PLANTS AND FUNGI IS A
CC HYDROGEN ION PUMP. THE PROTON GRADIENT IT GENERATES DRIVES THE
CC ACTIVE TRANSPORT OF NUTRIENTS BY H+-SYNPORT. THE RESULTING
CC EXTERNAL ACIDIFICATION AND/OR INTERNAL ALKALIZATION MAY MEDATE
CC GROWTH RESPONSES.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
CC -!- SURUNIT: POSSIBLY EXISTS AS AN HOMODIMER OR AN HOMOTRIMER.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- AS MANY AS 6 TO 8 CLOSELY RELATED GENES MAY ENCODE OTHER ISOFORMS
CC OF PLASMA MEMBRANE ATPASE IN TOMATO, LIKE THE LHA1 GENE PRODUCT
CC WHICH IS 96% IDENTICAL TO THE LHA2 GENE PRODUCT.
CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC (E1-E2 ATPASES).
DR PROSITE; PS00154; ATPASE_E1_E2.
KW HYDROLASE; HYDROGEN ION TRANSPORT; TRANSMEMBRANE; PHOSPHORYLATION;
KW ATP-BINDING; MULTIGENE FAMILY.
FT NON_TER 1 1
FT TRANSMEM <1 15 3 (POTENTIAL).
FT DOMAIN 16 27 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 28 50 4 (POTENTIAL).
FT DOMAIN 51 425 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 426 444 5 (POTENTIAL).
FT DOMAIN 445 460 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 461 484 6 (POTENTIAL).
FT DOMAIN 485 542 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 543 563 7 (POTENTIAL).
FT DOMAIN 562 578 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 579 595 8 (POTENTIAL).
FT DOMAIN 586 704 EXTRACELLULAR (POTENTIAL).
FT MOD_RES 81 81 PHOSPHORYLATION (BY SIMILARITY).
FT BINDING 175 175 ATP (BY SIMILARITY).
SQ SEQUENCE 704 AA; 78039 MW; C1CB2D8A CRC32;

Query Match 4.6%; Score 98; DB 7; Length 704;
Best Local Similarity 27.4%; Pred. No. 9.31e-01;
Matches 23; Conservative 19; Mismatches 38; Indels 4; Gaps 4;

Db 480 fwaayet-qffprvfgvstlqtatddfkklasalyqvstlsqalfvtrsrswsfver 538
QY 168 FYSGHSSFSMYCMLFVALYQARMKGDWARLLRPT-LQFGLVAVS-IYVGLSRVSDYKHH 225
Db 539 pglilvvalivaglvatliavvas 562
QY 226 WSDVLT-GLIQGALVAILVAVTVS 248

RESULT 15
ID CADA_BACFI STANDARD; PRT; 723 AA.
AC P30336.
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE PROBABLE CADMIUM-TRANSPORTING ATPASE (EC 3.6.1.-) (CADMIUM EFFLUX
DE ATPASE).
GN CADA.
OS BACILLUS FIRMS.
OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OF4;
RX MEDLINE; 92332419
RA IVEY D.M., GUFFANTI A.A., SHEN Z., KUDYAN N., KRULWICH T.A.;
RL J. BACTERIOL. 174:4878-4884(1992).
CC -!- FUNCTION: THIS ELECTRONEUTRAL ANTIporter EJECTS ONE CD(2+) WHILE

```

```

CC ACCUMULATING TWO PROTONS BY AN ENERGY-DEPENDENT EFFLUX MECHANISM.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CONTAINS 4
CC POTENTIAL TRANSMEMBRANE DOMAINS
CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC (E1-E2 ATPASES).
CC -!- SIMILARITY: CONTAINS A COPY OF THE HEAVY-METAL-ASSOCIATED (HMA)
CC DOMAIN.
DR EMBL; M90750; G143753; -.
DR PIR; D42707; D42707.
DR PROSITE; PS00154; ATPASE_E1_E2.
DR PROSITE; PS01047; HMA.
KW HYDROLASE; TRANSMEMBRANE; PHOSPHORYLATION; ATP-BINDING;
KW CADMIUM RESISTANCE; METAL-BINDING.
FT DOMAIN 18 47 HMA.
FT METAL 23 23 CADMIUM (POTENTIAL).
FT METAL 26 26 CADMIUM (POTENTIAL).
FT MOD_RES 412 412 PHOSPHORYLATION (PROBABLE).
SQ SEQUENCE 723 AA; 78207 MW; 65222CA5 CRC32;

Query Match 4.6%; Score 97; DB 2; Length 723;
Best Local Similarity 33.3%; Pred. No. 1.24e+00;
Matches 20; Conservative 18; Mismatches 16; Indels 6; Gaps 6;

Db 89 edtkedkvpfykxkxstl-lyasllitfgylssyngvge-eniv-ttllf-lasmfi-gqls 143
QY 159 ERVKEGRLSFYSGHSSFSMYCMLFVAL-YLQARMKGDWARLLRPTLQFGLVAVSIYVGLS 217

Search completed: Tue Nov 4 10:21:36 1997
Job time : 55 secs.

```

WIRELESS

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPSrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Thu Nov 6 11:57:37 1997; MasPar time 148.35 Seconds
Tabular output not generated. 958.178 Million cell updates/sec

Title: >US-08-842-827-5
Description: (1-1362) from US08842827.seq
Perfect Score: 1362
N.A. Sequence: 1 GGCGCAGCTCTGCAAAAGTT.....TTTAAAAAAA 1362
Comp: CCGCGTCGACGCTTTTCAA.....AAATTTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 142080 seqs, 52183452 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq28
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29

Statistics: Mean 9.460; Variance 6.196; scale 1.527

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Match	Length	ID
1	81	5.9	1047	2 Q10572
2	67	4.9	1047	2 Q10572
3	48	3.5	91	9 Q51746
4	45	3.3	91	9 Q51746
5	39	2.9	204	1 N81164
6	39	2.9	204	1 N81164
7	38	2.8	114	12 Q70467
8	36	2.6	114	12 Q70467
9	36	2.6	114	12 Q70468
10	36	2.6	114	12 Q70468
11	36	2.6	114	12 Q70465
12	35	2.6	114	12 Q70465
13	36	2.6	114	12 Q70468
14	35	2.6	114	12 Q70469
15	34	2.5	114	12 Q70470

C	16	34	2.5	114	12	Q70465	Generic DNA sequence	6.25e-05
C	17	34	2.5	114	12	Q70466	Generic DNA sequence	6.25e-05
C	18	31	2.3	114	12	Q70472	Generic DNA sequence	1.82e-03
C	19	31	2.3	114	12	Q70470	Generic DNA sequence	1.82e-03
C	20	32	2.3	3871	2	N71302	HSV-1 gB and surround	5.98e-04
C	21	30	2.2	114	12	Q70471	Generic DNA sequence	5.45e-03
C	22	30	2.2	114	12	Q70471	Generic DNA sequence	5.45e-03
C	23	29	2.1	67	24	T14322	Primer used in the la	1.61e-02
C	24	29	2.1	114	12	Q70473	Generic DNA sequence	1.61e-02
C	25	29	2.1	114	12	Q70473	Generic DNA sequence	1.61e-02
C	26	28	2.1	114	12	Q70472	Generic DNA sequence	4.70e-02
C	27	27	2.0	81	21	T13611	DC43 TSAR library gen	1.35e-01
C	28	27	2.0	82	21	T13610	DC43 TSAR library gen	1.35e-01
C	29	27	2.0	565	6	Q35072	HCV envelope region n	1.35e-01
C	30	26	1.9	36	2	Q11195	Ballast Constituent c	3.80e-01
C	31	26	1.9	39	7	Q51787	Mixed oligonucleotide	3.80e-01
C	32	26	1.9	74	21	T13613	DC43 TSAR library gen	3.80e-01
C	33	26	1.9	75	21	T13612	DC43 TSAR library gen	3.80e-01
C	34	26	1.9	82	21	T13610	DC43 TSAR library gen	3.80e-01
C	35	26	1.9	565	6	Q35072	HCV envelope region n	3.80e-01
C	36	26	1.9	3871	2	N71302	HSV-1 gB and surround	3.80e-01
C	37	25	1.8	39	7	Q51787	Mixed oligonucleotide	1.05e+00
C	38	25	1.8	75	21	T13612	DC43 TSAR library gen	1.05e+00
C	39	24	1.8	201	2	N70195	Streptomyces protease	2.85e+00
C	40	24	1.8	498	3	N50034	Sequence encoding new	2.85e+00
C	41	24	1.8	501	3	N50024	Sequence encoding new	2.85e+00
C	42	24	1.8	501	3	N50032	Sequence encoding new	2.85e+00
C	43	24	1.8	501	3	N50030	Sequence encoding new	2.85e+00
C	44	24	1.8	787	28	T51183	Human breast specific	2.85e+00
C	45	25	1.8	1511	23	T29032	Endoglucanase (58 kDa	1.05e+00

ALIGNMENTS

RESULT 1
ID Q10572 standard; DNA; 1047 BP.
AC Q10572;
DT 09-APR-1991 (first entry)
DE Human Natriuretic Peptide Receptor B.
KW NPBB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
KW hyperaldosteronism; glaucoma; guanyl cyclase.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= signal sequence
FT Protein 12
FT /label= mature NPBB
FT Domain 23..455
FT /label= extracellular domain
FT /note= "binds natriuretic peptides A,B and C]"
FT Domain 456..456
FT /label= transmembrane domain
FT Domain 479..1047
FT /label= cytoplasmic domain
FT /note= "GC and protien kinase activity"
FT Modified -site 24..26
FT /label= N-glycos -site
FT Modified -site 35..37
FT /label= N-glycos -site
FT Modified -site 161..163
FT /label= N-glycos -site
FT Modified -site 195..197
FT /label= N-glycos -site
FT Modified -site 244..246
FT /label= N-glycos -site
FT Modified -site 277..279
FT /label= N-glycos -site
FT Modified -site 349..351
FT /label= N-glycos -site
FT Modified -site 600..602
FT /label= N-glycos -site
PN WO9100292-A.
PD 10-JAN-1991.

```

PF 22-JUN-1990; U03586.
PR 23-JUN-1989; US-370673.
PA (GETH ) GENENTECH INC.
PI Chang M, Goeddel D, Lowe D;
DR WPI: 91-036711/05.
DR N-PSDB; Q10324.
PT Natriuretic protein receptor B - for diagnosis and treatment of
PT kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
PS Claim 3; Fig 1; 49pp; English.
CC The sequence was derived from the DNA encoding natriuretic peptide
CC receptor B, NPRB, having guanyl cyclase (GC) activity and protein
CC kinase activity. The DNA can be inserted into expression vectors
CC for the prodn. of the protein, opt. after being mutated to produce
CC NPRB analogues. The protein (or variants) can be used in treatment of
CC (114,952). The protein (or variants) can be used in treatment of
CC natriuretic peptide disorders, and also to isolate peptides using
CC affinity chromatography. Antibodies with affinity for NPRB can
CC also be prepd.
SQ Sequence 1047 BP; 87 A; 15 C; 83 G; 51 T;

Query Match 5.9%; Score 81; DB 2; Length 1047;
Best Local Similarity 8.7%; Pred. No. 4.20e-31;
Matches 83; Conservative 265; Mismatches 598; Indels 9; Gaps 8;

Db 21 garntntnavnnnnhsyawaivrvnavanavngannvdrnnvsnngnacsnyna 80
Cp 1062 GAACATCACTGGGATGGTCTTGTCTGTATACGGGAGACAGTCCGCTGTAGAGGCCA 1003
Db 81 nnsavdnknyhdndnnngngvnyasvasrnashwnnnnttagvasnshndhyrtnv 140
Cp 1002 T-CATGATCAAGGTAACTGCAGAGGGGCGGAGCGGCTCTCGCCAAAGTGAAG 944
Db 141 rtgsankngnvvtnhghnnwtaraannvndartdrhntnngnngnngsnvnh 200
Cp 943 CGGGCTCGAGGTATAGCAACCAATACAGATAGTATAGTGGAGAGGAGCGATGGCCA 884
Db 201 nvvarngngnnnathnrrangrvnyvcgnnmhnnnnnnnnnnntngdyvnyndvngn 260
Cp 883 GAGAAGAAGACTTCTCGGCTTCTGGACTTTCTGCTCATCACTCTGTCATCTGTAGTTC 824
Db 261 snragntatgrnwdnrtrnnanarntntvntznnnnnnnnnnnnnnnnnnnnnnrard 320
Cp 823 TGAATCTAGCCTTCAGAGCAGTGTATGCTGGCTGAATACAGGTTGCGAGACACTCAAGAAG 764
Db 321 ngvnnngsnmnnnagcnydgngnnvnnntnngtrndgrnrvnkmgrryhygtgv 380
Cp 763 TGAGGACGAGCGCCCTATGGACACTTTGGCAATGT-CTGTAAAGACTGGCTGATGGC 705
Db 381 vmckndrntdnvnmngdndsg-dhnnaaahysganknnwttgrnnnwkvgnnsdnnnc 439
Cp 704 ACAGCCAAAGAGAGAGCAGCCACTTGCITATAGAGTGTGCTGCCACCTAGGGGTTCGAAAT 645
Db 440 andhdnscktnstnaenvangtntnmgvssnnnnrknmmnnknnasmmwnrwnnn 499
Cp 644 CGTCGA-CCGCGACTTCTTCAGGATAATAGATCGGTAGAAATCCCGCTGATGATCGCGA 586
Db 500 ngnsnryhkggrtnsnrgssysnntahgkynnntnngnntghnkvnnkhvknrrnn 559
Cp 585 GGATGCAATGACATCCCGACCGGAC-AGAGCAGCAGCGTCAATTATGTGCTCACCAGTT 527
Db 560 ntrvnnnnkhmrdrnnnnhnrnngacndnnnnncvntnycnrgsnndnnndnnndwnn 619
Cp 526 TTCAGTGGGTACTGTGCTCTCATCTATTCAGTAAACCCCTCGGTGGTAAGCTTGATG 467
Db 620 ryennndvkgmannhnnssnshgsksnsvdsvrnnkntdygnasrntannddna 679
Cp 466 GTGCTTGTCTCGATGATAGAGAAAGAGGAGGCGCCGCTGAAGAGGAGAAG-AGGTCGAG 408
Db 680 nyakknntannngnnnttgmnaadvysngnnnnnnnnnnnnnnnnngnngndsnknvuk 739
Cp 407 GCAGATGAGCAGACCGCTGTGT-GCCCTCTCTCTGGGGTGTGTGTGACGCGCGG 350
Db 740 vrngnrynrsndrtnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 799

```

```

Cp 349 CTGCGCGCGTCTTCTCTCCGGGAGCATCGCTTTGTCTGCTACTTGTAGTTTGCATGGCG 290
Db 800 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 859
Cp 289 CTGGCTGCGCGCGAGCCTCCGCGCCCGGAGAGCTCCCGCAACAGCAGCAGCACACGCC 230
Db 860 vtynnsdnvngntansans-tmmnvvtnnndndntcndannndndvkvntngdaymvsagn 918
Cp 229 AGCGCCCGGGTGGCTCTCTGCGCCAGAGGCTGCTGCGGATAGTGGCGGCTCGGCCCGGCT 170
Db 919 ngngnrrhannnnarmanndavsnrrnrhrndhnnrnnrnnngvhtgncagvvgnk 973
Cp 169 CGGGGCGCGCGCTAGAGTGCGAGCGCGGGGCTGCTGCTCCAACTGCAGAAGGT 115

RESULT 2
ID Q10572 standard; DNA; 1047 BP.
AC Q10572;
DT 09-APR-1991 (first entry)
DE Human Natriuretic Peptide Receptor B.
KW NPRB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
KW hyperaldosteronism; glaucoma; guanyl cyclase.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= signal sequence
FT Protein 12
FT /label= mature NPRB
FT Domain 23..455
FT /label= extracellular domain
FT /note= "binds natriuretic peptides A,B and C]"
FT Domain 456..456
FT /label= transmembrane domain
FT Domain 479..1047
FT /label= cytoplasmic domain
FT /note= "GC and protien kinase activity"
FT Modified-site 24..26
FT /label= N-glycos_site
FT Modified-site 35..37
FT /label= N-glycos_site
FT Modified-site 161..163
FT /label= N-glycos_site
FT Modified-site 195..197
FT /label= N-glycos_site
FT Modified-site 244..246
FT /label= N-glycos_site
FT Modified-site 277..279
FT /label= N-glycos_site
FT Modified-site 349..351
FT /label= N-glycos_site
FT Modified-site 600..602
FT /label= N-glycos_site
FT WO9100292-A.
PN 10-JAN-1991.
PE 22-JUN-1990; U03586.
PR 23-JUN-1989; US-370673.
PA (GETH ) GENENTECH INC.
PI Chang M, Goeddel D, Lowe D;
DR WPI: 91-036711/05.
DR N-PSDB; Q10324.
PT Natriuretic protein receptor B - for diagnosis and treatment of
PT kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
PS Claim 3; Fig 1; 49pp; English.
CC The sequence was derived from the DNA encoding natriuretic peptide
CC receptor B, NPRB, having guanyl cyclase (GC) activity and protein
CC kinase activity. The DNA can be inserted into expression vectors
CC for the prodn. of the protein, opt. after being mutated to produce
CC NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr-
CC 114,952). The protein (or variants) can be used in treatment of
CC natriuretic peptide disorders, and also to isolate peptides using
CC affinity chromatography. Antibodies with affinity for NPRB can
CC also be prepd.

```



```

FT /*tag= a
FT /function= multiple cloning site
FT primer_bind 187..204
FT /*tag= b
FT /*tag= c
PN EP-285123-A.
PD 05-MAY-1988. 105163.
PR 30-MAR-1988; US-034819.
PR 03-APR-1987; US-034819.
PA (SUO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prep of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; p; English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 2.9%; Score 39; DB 1; Length 204;
Best Local Similarity 18.1%; Pred. No. 1.83e-07;
Matches 29; Conservative 68; Mismatches 61; Indels 2; Gaps 2;

Db 29 aytccbcagcgagcbcyrraggycccggywccgagcycaaycdchvccgymr 88
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
929 ATCTCTCAGCGCCGCTTCTACTTGGCAGAGAGCCCGCTGCTCGGCC-CTC-CTGCAG 986
Db 89 ttthyrmrbyrvdyrnsdaaawccyrsvkydcynachdhdyvbbvynvhn 148
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
987 TTCACCTTGATGATGATGCGCTTCTACACGGAGCTGCTCGGTATCAGACACAGCAC 1046
Db 149 nnncccnhvhvbnhnrwayrvhddvhhccv 188
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1047 CATCCCATGATGTTCTGGCAGGATTTGCTCAAGAGGCC 1086

RESULT 6
ID NB1164 standard; DNA; 204 BP.
AC NB1164;
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KW E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
FH Key Location/Qualifiers
FT misc_feature 19..69
FT /*tag= a
FT /function= multiple cloning site
FT primer_bind 187..204
FT /*tag= b
PN EP-285123-A.
PD 05-MAY-1988. 105163.
PR 30-MAR-1988; US-034819.
PR 03-APR-1987; US-034819.
PA (SUO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prep of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; p; English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to

```

```

CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 2.9%; Score 39; DB 1; Length 204;
Best Local Similarity 16.8%; Pred. No. 1.83e-07;
Matches 25; Conservative 64; Mismatches 59; Indels 1; Gaps 1;

Db 47 yrraggycccggywccgagcycaaycdchvccgymtttthyrmrbyrvdyn 106
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
576 TGACGATCCCCACGGCAGACAGACAGCGCTCATTTATTGCTCACCAGTTTTCAGTGGGT 517
Db 107 rsdaaawccyrsvkydcynachdhdyvbbvynvhnncnccbnhvhcnvhhb 166
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
516 ACTTGATGCTCT-CATCATGTCAGTAAACCCCTCGGTGTAAGGCTGATGCTGCTGTC 458
Db 167 nhnrwayrvhddvhhccvchccgat 195
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
457 TCGATGATGAGGAAGGGAGGCCGCCAT 429

RESULT 7
ID Q70467 standard; DNA; 114 BP.
AC Q70467;
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60
FT /*tag= a
FT /note= "this sequence represents '2'; 2 can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN WO9418318-A.
PD 18-AUG-1994. U00977.
PR 01-FEB-1994; US-013416.
PR 01-FEB-1993; US-176500.
PR 30-DEC-1993; US-189331.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI: 94-279739/34.
DR P-PSDB; R65153.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 35; 255pp; English.
CC Q70467 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)16(TGC)(NNB)12(NNB)16(TGC)(NNB)1X. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in Q70466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active. They may further comprise a linker
CC peptide between the 2 domains. The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or compens. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,

```



```

QY 326 CCGGAGAGCAAGACGGCGCGCTCAACAACAACCGGAGGAGCGGCAG 385
Db 63 bnnbnnbnnbnnbnnbgtcnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
QY 386 CAAGCGGGTGTCTCATCTGCTGACCTCTTCTGCTTCATGCGGGC 437

RESULT 10
ID Q70466 standard; DNA; 114 BP.
AC Q70466;
DE Generic DNA sequence to generate a random TSAR-9 peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60
FT /tag= a
FT /note= "this sequence represents '2'; 2 can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYN-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI; 94-279739/34.
DR P-PSDB; R65152.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 35; 255pp; English.
CC Q70466 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)1(TGC)(NNB)10(TGC)2(NNB)42(NNB)8(TGC)(NNB)
CC -9Y. X and Y are flanking restriction sites (X is not the same as Y)
CC that are not specified further. Other generic sequences are shown in
CC Q70466-68. Other specific peptides generated by these generic sequences
CC are shown in R65151-54. TSARs are concatenated heterofunctional proteins
CC or peptides, comprising at least two functional regions - a binding
CC domain with affinity for a ligand and a second effector peptide portion
CC that is chemically or biologically active. They may further comprise a
CC linker peptide between the 2 domains. The oligonucleotides are also
CC designed so that the expressed peptide contains 2 or 4 cysteine residues
CC positioned in, or flanking, the unpredicted or variant residues. These
CC residues confer some degree of conformational rigidity to the peptides.
CC The TSARs or compns. comprising a TSAR binding domain can be used in
CC vivo to deliver a chemically or biologically active moiety, eg. metal
CC ion, radioisotope, peptide, toxin or enzyme, to the specific target or
CC on the cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need for
CC complex methods of hybridoma formation or in vivo antibody production.
CC The TSARs are easily characterised and have designed activity allowing
CC direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 4 C; 4 G; 4 T;

Query Match 2.6%; Score 36; DB 12; Length 114;
Best Local Similarity 7.3%; Pred. NO. 6.24e-06;
Matches 8; Conservative 31; Mismatches 71; Indels 0; Gaps 0;

Db 5 gcnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 64
QY 346 GCACCGCGGCTCAACAACAACCGGAGGAGCGGCTGCTCATCT 405
Db 65 nbnnbnnbnnbnnbnnbgtcnnbnnbnnbnnbnnbnnbnnbnnbnn 114
QY 406 GCCTCGACCTCTTCTGCTCTTCATGCGGGCTCCCTTCCTCATC 455

```

```

RESULT 11
ID Q70465 standard; DNA; 114 BP.
AC Q70465;
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60
FT /tag= a
FT /note= "this sequence represents '2'; 2 can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYN-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI; 94-279739/34.
DR P-PSDB; R65150 and R65151.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 35; 255pp; English.
CC Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)6(TGC)(NNB)11z(NNB)14(TGC)(NNB)3Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in Q70466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active. They may further comprise a linker
CC peptide between the 2 domains. The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or compns. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need
CC for complex methods of hybridoma formation or in vivo antibody
CC production. The TSARs are easily characterised and have designed
CC activity allowing direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 2.6%; Score 36; DB 12; Length 114;
Best Local Similarity 3.6%; Pred. NO. 6.24e-06;
Matches 4; Conservative 34; Mismatches 74; Indels 0; Gaps 0;

Db 3 bnnbnnbnnbnnbnnbgtcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62
QY 326 CCGGAGAGCAAGACGGCGCGCTCAACAACAACCGGAGGAGCGGCAG 385
Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
QY 386 CAAGCGGGTGTCTCATCTGCTGACCTCTTCTGCTTCATGCGGGC 437

RESULT 12
ID Q70469 standard; DNA; 114 BP.
AC Q70469;
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;

```

KW direct; rapid; detection; screening; treatment; generic; ss.
 OS Synthetic.
 FH Key
 FT misc_feature 55..60
 FT Location/Qualifiers
 FT /tag= a
 FT /note= "this sequence represents '2'; 2 can be a
 FT sequence of 6, 9 or 12 nucleotides (see
 FT comments)"
 PN WO9418318-A.
 PD 18-AUG-1994.
 PF 01-FEB-1994; US-013416.
 PR 01-FEB-1994; US-013416.
 PR 30-DEC-1993; US-176500.
 PR 31-JAN-1994; US-189331.
 PA (UINC-) UNIV NORTH CAROLINA.
 PI Fowlkes DM, Kay BK;
 DR WPI: 94-279739/34.
 DR P-PSDB: R65154.
 PT Identifying proteins or peptide(s) which bind a ligand - by
 PT screening a recombinant vector library expressing fusion proteins
 PT comprising a binding domain and an effector domain
 PS Disclosure; Page 35; 255pp; English.
 CC 070469 is a generic DNA sequence used to generate random TSAR peptide
 CC (TGC)(NNE)62(NNE)2(TGC)(NNE)14(TGC)Y. X and Y are flanking restriction
 CC sites (X is not the same as Y) that are not specified further. This
 CC sequence generates peptides that are cloverleaf in structure. Other
 CC generic sequences are shown in 070465-68. Other specific peptides
 CC generated by these generic sequences are shown in 070465-68. TSARs are
 CC concatenated heterofunctional proteins or peptides, comprising at least
 CC two functional regions - a binding domain with affinity for a ligand and
 CC a second effector peptide portion that is chemically or biologically
 CC active. They may further comprise a linker peptide between the 2 domains.
 CC The oligonucleotides are also designed so that the expressed peptide
 CC contains 2 or 4 cysteine residues positioned in, or flanking, the
 CC unpredicted or variant residues. These residues confer some degree of
 CC conformational rigidity to the peptides. The TSARs or compens. comprising
 CC a TSAR binding domain can be used in vivo to deliver a chemically or
 CC biologically active moiety, eg. metal ion, radioisotope, peptide, toxin
 CC or enzyme, to the specific target or on the cell. They can also replace
 CC the function of macromolecules, eg. monoclonal or polyclonal antibodies
 CC and therefore circumvent the need for complex methods of hybridoma
 CC formation or in vivo antibody production. The TSARs are easily
 CC characterised and have designed activity allowing direct and rapid
 CC detection in a screening process.
 SQ Sequence 114 BP; 0 A; 4 C; 4 G; 4 T;
 Query Match 2.6%; Score 35; DB 12; Length 114;
 Best Local Similarity 4.7%; Pred. No. 1.99e-05;
 Matches 5; Conservative 32; Mismatches 70; Indels 0; Gaps 0;
 Db 6 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 65
 QY 320 GATCGTCCGGAGAGCAAGACGGCGGCGGCTCAACACCCGAGGAGGAG 379
 Db 66 btgcnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 112
 QY 380 CGGACGAAGCGGTGTGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCT 426
 RESULT 13
 ID 070468 standard; DNA; 114 BP.
 AC 070468;
 DT 05-APR-1995 (first entry)
 DE Generic DNA sequence to generate a random TSAR peptide library.
 KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
 KW effector domain; concatenated heterofunctional protein; linker;
 KW direct; rapid; detection; screening; treatment; generic; ss.
 OS Synthetic.
 FH Key
 FT misc_feature 55..60
 FT Location/Qualifiers
 FT /tag= a
 FT /note= "this sequence represents '2'; 2 can be a
 FT sequence of 6, 9 or 12 nucleotides (see

FT comments)"
 PN WO9418318-A.
 PD 18-AUG-1994.
 PF 01-FEB-1994; US-013416.
 PR 01-FEB-1994; US-013416.
 PR 30-DEC-1993; US-176500.
 PR 31-JAN-1994; US-189331.
 PA (UINC-) UNIV NORTH CAROLINA.
 PI Fowlkes DM, Kay BK;
 DR WPI: 94-279739/34.
 DR P-PSDB: R65154.
 PT Identifying proteins or peptide(s) which bind a ligand - by
 PT screening a recombinant vector library expressing fusion proteins
 PT comprising a binding domain and an effector domain
 PS Disclosure; Page 35; 255pp; English.
 CC 070468 is a generic DNA sequence used to generate random TSAR (Totally
 CC represented as follows: X(NNE)11(TGC)(NNE)62(NNE)10(Y). X
 CC and Y are flanking restriction sites (X is not the same as Y) that are
 CC not specified further. Other generic sequences are shown in 070465-68.
 CC Other specific peptides generated by these generic sequences are shown in
 CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
 CC comprising at least two functional regions - a binding domain with
 CC affinity for a ligand and a second effector peptide portion that is
 CC chemically or biologically active. They may further comprise a linker
 CC peptide between the 2 domains. The oligonucleotides are also designed so
 CC that the expressed peptide contains 2 or 4 cysteine residues positioned
 CC in, or flanking, the unpredicted or variant residues. These residues
 CC confer some degree of conformational rigidity to the peptides. The TSARs
 CC or compens. comprising a TSAR binding domain can be used in vivo to
 CC deliver a chemically or biologically active moiety, eg. metal ion,
 CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
 CC cell. They can also replace the function of macromolecules, eg.
 CC monoclonal or polyclonal antibodies and therefore circumvent the need
 CC for complex methods of hybridoma formation or in vivo antibody
 CC production. The TSARs are easily characterised and have designed activity
 CC allowing direct and rapid detection in a screening process.
 SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;
 Query Match 2.6%; Score 36; DB 12; Length 114;
 Best Local Similarity 3.6%; Pred. No. 6.24e-06;
 Matches 4; Conservative 34; Mismatches 74; Indels 0; Gaps 0;
 Db 3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 62
 Cp 389 CTGTGCTCCGCTCTCTCTCGGTTGTTGTTGAGCGCGCGGCTCTCTCTCTC 330
 Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 114
 Cp 329 CGGACGATCGCTTTGCTGCTACTGTAGTTTTCATGCGCTGCTGCGCG 278
 RESULT 14
 ID 070469 standard; DNA; 114 BP.
 AC 070469;
 DT 07-APR-1995 (first entry)
 DE Generic DNA sequence to generate a random TSAR peptide library.
 KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
 KW effector domain; concatenated heterofunctional protein; linker;
 KW direct; rapid; detection; screening; treatment; generic; ss.
 OS Synthetic.
 FH Key
 FT misc_feature 55..60
 FT Location/Qualifiers
 FT /tag= a
 FT /note= "this sequence represents '2'; 2 can be a
 FT sequence of 6, 9 or 12 nucleotides (see
 FT comments)"
 PN WO9418318-A.
 PD 18-AUG-1994.
 PF 01-FEB-1994; US-013416.
 PR 01-FEB-1994; US-013416.
 PR 30-DEC-1993; US-176500.
 PR 31-JAN-1994; US-189331.

PA (UUNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI: 94-279739/34.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure: Page 35; 25pp; English.
CC Q70469 is a generic DNA sequence used to generate random TSAR peptide
CC This generic formula can be represented as follows: X(TGC)(NNB)10-
CC (TGC)(NNB)62(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking restriction
CC sites (X is not the same as Y) that are not specified further. This
CC sequence generates peptides that are cloverleaf in structure. Other
CC generic sequences are shown in Q70465-68. Other specific peptides
CC generated by these generic sequences are shown in R65150-54. TSARS are
CC concatenated heterofunctional proteins or peptides, comprising at least
CC two functional regions - a binding domain with affinity for a ligand and
CC a second effector peptide portion that is chemically or biologically
CC active. They may further comprise a linker peptide between the 2 domains.
CC The oligonucleotides are also designed so that the expressed peptide
CC contains 2 or 4 cysteine residues positioned in, or flanking, the
CC conformational rigidity to the peptides. The TSARS or compsns. comprising
CC a TSAR binding domain can be used in vivo to deliver a chemically or
CC biologically active moiety, eg. metal ion, radioisotope, peptide, toxin
CC or enzyme, to the specific target or on the cell. They can also replace
CC the function of macromolecules, eg. monoclonal or polyclonal antibodies
CC and therefore circumvent the need for complex methods of hybridoma
CC formation or in vivo antibody production. The TSARS are easily
CC characterised and have designed activity allowing direct and rapid
CC detection in a screening process.
SQ Sequence 114 BP; 0 A; 4 C; 4 G; 4 T;

Query Match 2.6%; Score 35; DB 12; Length 114;
Best Local Similarity 4.7%; Pred. No. 1.99e-05;
Matches 5; Conservative 32; Mismatches 70; Indels 0; Gaps 0;

Db 6 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 65
Cp 419 GAAGAGCGGGCTGCCCGCTCTTGTCTCTCCGGGACGATCGCTTTGT 313
Db 66 btgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 112
Cp 359 GAGCGCGGGCTGCCCGCTCTTGTCTCTCCGGGACGATCGCTTTGT 313

RESULT 15
ID Q70470 standard; DNA; 114 BP.
AC Q70470.
DT 10-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60
FT /tag= a
FT /note= "encoded by z (see comments)"
PN W09418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994. 000977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UUNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI: 94-279739/34.
P-PSDB; R58378.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure: Page 36; 25pp; English.
CC Q70470 is a generic DNA sequence used to generate random TSAR (Totally

CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)4(CAC)(NNB)4(CAC)(NNB)82(NNB)6(CAC)(NNB)8
CC -(CAC)2(NNB)Y. X and Y are flanking restriction sites (X is not the same
CC as Y) that are not specified further. The peptides generated by this and
CC other generic sequences (Q70471-73) have invariant histidine residues
CC incorporated into variant sequences. TSARS are concatenated
CC heterofunctional proteins or peptides, comprising at least two functional
CC regions - a binding domain with affinity for a ligand and a second
CC effector peptide portion that is chemically or biologically active. They
CC may further comprise a linker peptide between the 2 domains. The TSARS
CC or compsns. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need
CC for complex methods of hybridoma formation or in vivo antibody
CC production. The TSARS are easily characterised and have designed
CC activity allowing direct and rapid detection in a screening process.
SQ Sequence 114 BP; 5 A; 10 C; 0 G; 0 T;

Query Match 2.5%; Score 34; DB 12; Length 114;
Best Local Similarity 8.0%; Pred. No. 6.25e-05;
Matches 9; Conservative 31; Mismatches 72; Indels 0; Gaps 0;

Db 3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 62
Qy 335 CAAGAAGCGGCGCAGCCCGCGCTCAACAACACCCGAGGAGCGCGCAGCAACGCGGT 394
Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 114
Qy 395 GCTGCTCATCTGCTCGACCTCTTCTGCTCTTATGCGGCGCTCCCTTC 446

Search completed: Thu Nov 6 12:00:32 1997
Job time : 175 secs.

WATERMAN

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_nnn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Thu Nov 6 12:00:58 1997; MasPar time 719.25 Seconds
Tabular output not generated. 972.045 Million cell updates/sec

Title: >US-08-842-827-5
Description: (1-1362) from US08842827.seq
Perfect Score: 1362
N.A. Sequence: 1 GCGCGAGCTCTGCAAAAGTT.....TTTAAAAAAAAAAAAAAA 1362
Comp: CCGCGTCGAGAGCTTTTCA.....AAATTTTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 707517 seqs, 256659390 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: EST-STS
1:EST1 2:EST2 3:EST3 4:EST4 5:EST5 6:EST6 7:EST7 8:EST8
9:EST9 10:EST10 11:EST11 12:EST12 13:EST13 14:EST14
15:EST15 16:EST16 17:EST17 18:EST18 19:EST19 20:EST20
21:EST21 22:EST22 23:EST23 24:EST24 25:EST25 26:EST26
27:EST27 28:EST28 29:EST29 30:EST30 31:EST31 32:EST32
33:EST33 34:EST34 35:EST35 36:EST36 37:EST37 38:EST38
39:EST39 40:EST40 41:EST41 42:EST42 43:EST43 44:EST44
45:EST45 46:EST46 47:EST47 48:EST48 49:EST49 50:EST50
51:EST51 52:EST52 53:EST53 54:EST54 55:EST55 56:EST56
57:EST57 58:EST58 59:EST59 60:EST60 61:EST61 62:EST62
63:EST63 64:EST64 65:EST65 66:EST66 67:EST67 68:EST68
69:EST69 70:EST70 71:EST71 72:EST72 73:EST73 74:EST74
75:EST75 76:EST76 77:EST77 78:EST78 79:EST79 80:EST80
81:EST81 82:EST82 83:EST83 84:EST84 85:EST85 86:EST86
87:EST87 88:EST88 89:EST89 90:EST90 91:EST91 92:EST92
93:EST93 94:EST94 95:EST95 96:EST96 97:EST97 98:EST98
99:EST99

EST-STS-TWO
100:EST100 101:EST101 102:EST102 103:EST103 104:EST104
105:EST105 106:EST106 107:EST107 108:EST108 109:EST109
110:EST110 111:EST111 112:EST112 113:EST113 114:EST114
115:EST115 116:EST116 117:EST117 118:EST118 119:EST119
120:EST120 121:EST121 122:EST122 123:EST123 124:EST124
125:EST125 126:EST126 127:EST127 128:EST128 129:EST129
130:EST130 131:EST131 132:EST132 133:EST133 134:EST134
135:EST135 136:EST136 137:EST137 138:EST138 139:EST139
140:EST140 141:EST141 142:EST142 143:EST143 144:EST144
145:EST145 146:EST146 147:EST147 148:EST148 149:EST149
150:EST150 151:EST151 152:EST152 153:EST153 154:EST154
155:EST155 156:EST156 157:EST157 158:EST158 159:EST159
160:EST160 161:EST161 162:EST162 163:EST163 164:EST164
165:EST165 166:EST166 167:EST167 168:EST168 169:EST169
170:EST170 171:EST171 172:EST172 173:EST173 174:EST174

175:EST175 176:EST176 177:EST177 178:EST178 179:EST179
180:EST180 181:EST181 182:EST182 183:EST183 184:EST184
185:EST185 186:EST186 187:EST187 188:EST188 189:EST189
190:EST190 191:EST191 192:EST192 193:EST193 194:EST194
195:EST195 196:EST196 197:EST197 198:EST198

Statistics: Mean 11.544; Variance 2.649; scale 4.357

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
c 1	512	37.6	555	161	AA056693	z81b05.s1 Soares pre	0.00e+00
c 2	482	35.4	572	172	AA058383	z81b05.r1 Soares pre	0.00e+00
c 3	478	35.1	566	185	AA135741	z111f05.s1 Soares pre	0.00e+00
c 4	440	32.3	449	173	W70040	z49b06.s1 Soares fet	0.00e+00
c 5	405	29.7	451	83	H50533	yp08a11.s1 Homo sapie	0.00e+00
c 6	402	29.5	480	54	H04639	yj49a08.s1 Homo sapie	0.00e+00
c 7	401	29.4	428	191	AA156340	z054f04.s1 Strata gene	0.00e+00
c 8	388	28.5	413	102	N68923	z469f12.s1 Homo sapie	0.00e+00
c 9	380	27.9	402	103	N70674	z432a04.s1 Homo sapie	0.00e+00
c 10	376	27.6	446	29	R63796	y115h04.s1 Homo sapie	0.00e+00
c 11	357	26.2	431	29	R63741	y115h04.s1 Homo sapie	0.00e+00
c 12	353	25.9	463	43	H02469	yj35d09.s1 Homo sapie	0.00e+00
c 13	327	24.0	348	54	H04659	yj49a08.r1 Homo sapie	0.00e+00
c 14	309	22.7	475	43	H02468	yj35d09.r1 Homo sapie	0.00e+00
c 15	298	21.9	469	117	W31303	z64b01.r1 Soares fet	0.00e+00
c 16	297	21.8	370	30	R71019	y150a11.r1 Homo sapie	0.00e+00
c 17	291	21.4	442	60	H12563	yj12d10.s1 Homo sapie	0.00e+00
c 18	287	21.1	422	69	H75437	yr69g02.s1 Homo sapie	0.00e+00
c 19	283	20.8	402	30	R71020	y150a11.s1 Homo sapie	0.00e+00
c 20	273	20.0	395	10	T92854	ye27d03.r1 Homo sapie	0.00e+00
c 21	269	19.8	279	69	H77957	ys10h08.s1 Homo sapie	0.00e+00
c 22	269	19.8	398	138	AA048371	m128d11.r1 Soares mou	0.00e+00
c 23	267	19.6	485	60	H12562	yj12d10.r1 Homo sapie	0.00e+00
c 24	259	19.0	275	64	H26102	y144e12.s1 Homo sapie	0.00e+00
c 25	259	19.0	502	113	W00564	z469f12.r1 Homo sapie	0.00e+00
c 26	253	18.6	292	20	R00691	ye74a10.s1 Homo sapie	0.00e+00
c 27	254	18.6	344	69	H07856	ys09h08.s1 Homo sapie	0.00e+00
c 28	249	18.3	332	40	R21805	yh21g05.r1 Homo sapie	0.00e+00
c 29	243	17.8	546	5	T71976	yc63g02.s1 Homo sapie	0.00e+00
c 30	235	17.3	278	185	AA135784	z111f05.r1 Soares pre	0.00e+00
c 31	227	16.7	241	40	R21755	yh21g05.s1 Homo sapie	0.00e+00
c 32	227	16.7	246	181	AA092191	114679.seq.F Fetal he	0.00e+00
c 33	228	16.7	346	122	W61641	m835a11.r1 Life Tech	0.00e+00
c 34	223	16.4	225	109	HSC3MH032	H. sapiens partial CD	0.00e+00
c 35	217	15.9	223	110	HUM336804B	Human aorta cDNA 5'-e	0.00e+00
c 36	207	15.2	292	20	R00690	ye74a10.r1 Homo sapie	0.00e+00
c 37	204	15.0	263	145	R75377	MB05338R Mouse brain,	1.63e-298
c 38	202	14.8	203	165	C18274	Human placenta cDNA 5	6.13e-295
c 39	200	14.7	314	145	R75376	MB05338 Mouse brain,	2.29e-291
c 40	188	13.8	208	165	C16881	Human placenta cDNA 5	5.56e-270
c 41	159	11.7	372	39	R19424	yg25b02.r1 Homo sapie	1.10e-218
c 42	153	11.2	338	109	HSC3MH031	H. sapiens partial CD	3.76e-208
c 43	138	10.1	402	6	T78825	yc96f03.r1 Homo sapie	5.69e-182
c 44	133	9.8	440	163	AA105463	mm92b06.r1 Strata gene	2.68e-173
c 45	113	8.3	214	191	AA156375	z054f04.r1 Strata gene	6.24e-139

ALIGNMENTS

1
RESULT
LOCUS
DEFINITION
ACCESSION
NID
KEYWORDS

AA056693 555 bp mRNA EST 17-SEP-1996
z81b05.s1 Soares pregnant uterus Ndhpu Homo sapiens cDNA clone
489201 3' similar to WP:128D9.3 CE02068 ;
AA056693
g1549094
EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 553)
REFERENCE Hallier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennan, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 476.
FEATURES
source
1..553
/organism="Homo sapiens"
/note="Organ: uterus; Vector: p7T73-Pac; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'
AACTGGAAGAAATTCGGCGCGCTTTTTTTTTTTTTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."
/clone_lib="Soares pregnant uterus NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
complement(<1..>553)
BASE COUNT 146 a 121 c 159 g 129 t
ORIGIN
Query Match 37.68; Score 512; DB 161; Length 553;
Best Local Similarity 98.9%; Pred. No. 0.00e+00;
Matches 548; Conservative 0; Mismatches 0; Indels 6; Gaps 6;
Db 1 aaaaatcagtcgagcaaaagtgtttccctacattctactctgatgatgagagca 60
Cp 1346 AAAAATCAGTCGGCAAAAGTTTTCCTACATTCATCTGTCATGAGATTGGAGCA 1287
Db 61 gcaagaactgtctgcagcagtcattttacaaaacagctcaggagtggtggcaccta 120
Cp 1286 GCAAGAACTTGTCTGACAGTCATTTTACAAAACAGCTCAGAGGTGGTGGCACCTA 1227
Db 121 catcatgttgggtgattgtctctgccaataatgtccacaggtgaaggattctctccg 180
Cp 1226 CATCATGTTGTGTAATGTGTCTGTCAATATGTCACACAGGTGAAGGATTTCCTCCG 1167
Db 181 gatagcaggggcaggcaggagcgctcttagtcttaaaaggctcacacagaaga 240
Cp 1166 GATAGCAGGGCAGGAGGAGAGCGTCTTGTAGTCTTGAAGAGGTGACACAGAGAA 1107
Db 241 aactatgcagcagccaccaggggtccttgagcaaatcctgccagaacatcactgggatg 300
Cp 1106 AACTATCAGACGCCACCAGGGTCTCTTGAGCAATCTCTGCCAGACATCACTGGGATG 1047
Db 301 gtgccttggtctgatacgcgagacagtcctccgtgtagaagggccatcatcatcaagtga 360
Cp 1046 GTGCTTGTGCTCTGATACGGGAGACAGTCCTCCGTGTAGAAGGCCATCATGATCAAGGTG-A 988

Db 361 acttgacagggggccggagcagggcggtcctcccaagtgaagcggcggtcgaggta 420
Cp 987 ACT-GCAGGAGGGCCGGAGCAGCGGGGTCTCGCCAAGT-GAAGCGGCTCGAGGTA 930
Db 421 tagcaccacaatcacagcatagtgatcatgagagagggcagcagagaagaagactt 480
Cp 929 TAGCACCACAATCAGCATAGTGTACATGAGAGAGGAGGCATGCCAGAGAGAAGGACTT 870
Db 481 cctgggttctcctgagacttttgcgtcatcaccctctgcattgtaagtctgaatgta-cct 539
Cp 869 CTTGGCTTCTCGACTTT-GCTGTCTATCACCCTCTGCATCTGTA-GTTCTGAATGTAGCCT 812
Db 540 tcagagcagttgat 553
Cp 811 TCAGACCACTTGAT 798
RESULT 2
LOCUS AA058383 572 bp mRNA EST 19-SEP-1996
DEFINITION zklb05_r1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 489201 5' similar to TR.G1161100 G1161100 HYDROGEN PEROXIDE-INDUCIBLE PROTEIN ;
ACCESSION AA058383
NID g1551210
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 572)
REFERENCE Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennan, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 439.
FEATURES
source
1..572
/organism="Homo sapiens"
/note="Organ: uterus; Vector: p7T73-Pac; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'
AACTGGAAGAAATTCGGCGCGCTTTTTTTTTTTTTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."
/clone_lib="Soares pregnant uterus NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
complement(<1..>572)
BASE COUNT 130 a 160 c 137 g 141 t 4 others
ORIGIN

Query Match 35.4%; Score 482; DB 172; Length 572;
Best Local Similarity 98.0%; Pred. No. 0.00e+00;
Matches 542; Conservative 0; Mismatches 1; Indels 10; Gaps 8;

Db 1 gagaaagcaccataagccttaccacaggggttttactgcaatgatgagagcatcaag 60
 QY 456 GAGACAAGCACCATCAAGCCTTACACCGAGGGTTTACTGCAATGATGAGAGCATCAAG 515
 Db 61 taccactgaaactggtgagacaataaataagcgtgtgtctctgtgcgtggatcgtc 120
 QY 516 TACCACCTGAAACTGTGTGACACAATAAATGACGTGTGTCTGTGCGTGGGATGTC 575
 Db 121 attgccatctcgcgatcatcacgggggaattctaccggatctctattaccctgaagaatcg 180
 QY 576 ATTGCCATCTCTGCGATCATCACGGGGGAATCTACCGGATCTATTACCTGAAGAATCG 635
 Db 181 cggtagcagattcagaacccctagtgagcagcactctataagcaagtggtgtctctc 240
 QY 636 CGGTGCGAGGATTCAGAACCCCTACGTGCGAGCACTCTATAAGCAAGTGGGTCTCTC 695
 Db 241 ttgggtgtgc-atcagccagtctttcacagacattgccaagtgtccatggcgctg 299
 QY 696 TTGGCTGTGCCATCAGCCAGTCTTTTCACAGACATTGCCAAAGTGTCCATTAGGGCGCTG 755
 Db 300 cgtctcaactcttgagtgctgcaaccctgatcttcagccagatcaactgctctgaaggc 359
 QY 756 CGTCTCACTTCTTGAGTGTCTGCAACCCTGATTTCAGCCAGATCAACTGCTCTGAAGGC 815
 Db 360 tacattcagaactcagatgcagaggtgatcacagcaagtcctcagggaagccaggaagtc 419
 QY 816 TACATTCAAGAACTACAGATGCAGAGGTGATGACAGCAAAAGTCCAGGAGCCAGGAAGTCC 875
 Db 420 ttctctctgcccagtcct 479
 QY 876 TCTCTCTGCCCCATGCT 933
 Db 480 tgcagggccgtttcactcttgagcagggccgctgctcnggcccctccctcctgag 539
 QY 934 TGCAGGCCCGCTT-CACAT-GGCGAGGAGCCCG-CTCG-CTCC--GGCCCCCTCC-TGCAG 596
 Db 540 ttcaccttgatca 552
 QY 987 TTCACCTTGATCA 999

RESULT 3
 LOCUS A135741 566 bp mRNA EST 30-NOV-1996
 DEFINITION z11f05.s1 Soares pregnant uterus NBHPU Homo sapiens cDNA clone 501633 3'.
 ACCESSION A135741
 NID 91596753
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Euthera; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 566)
 REFERENCE Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevas, K., E., Waterston, R., Williamson, A., Wohlmann, P., and Wilson, R.
 The WashU-Merck EST Project
 Unpublished (1995)

Contact: Wilton RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available royalty-free through LNL ; contact the IMAGE Consortium (infoimage.lnl.gov) for further information.
 Seq primer: -40M13 fwd. from Amersham
 High quality sequence stop: 393.

FEATURES
 source 1..566
 Location/Qualifiers
 /organism="Homo sapiens"
 /note="Organ: uterus; Vector: p7T3-Pac; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'
 AACTGGAGAATTCGCGCGCGCTTTTTTTTTTTT 3']
 double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."
 /clone="501633"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /lab_lib="Soares pregnant uterus NBHPU"
 complement(<1..>566)
 BASE COUNT 149 a 126 c 163 g 126 t 2 others
 ORIGIN

Query Match 35.18; Score 478; DB 185; Length 566;
 Best Local Similarity 97.5%; Pred. No. 0.00e+00;
 Matches 551; Conservative 0; Mismatches 2; Indels 12; Gaps 12;

Db 1 aaaaatcagtcgggcaaaagtgtttccctacattcttcaaaaaacagctcaggaggtgggtggcaccta 60
 Cp 1346 AAAAATCAGTCGCGGCAAAAGTTTTTCCTACATCTCTGATGAGATTTGGAGACA 1287
 Db 61 gcaagaacttgcgtcagcagtcattttcaaaaaacagctcaggaggtgggtggcaccta 120
 Cp 1286 GCAAGAATCTGCTGACGAGTCATTTTACAAAAACAGCTCAGGAGTGGTGGCACCTA 1227
 Db 121 catcatattggtgattgttctctgtaataatgtccacaggtgaaaggattctctccg 180
 Cp 1226 CATCATTTTGTGTGATTTGCTCTCAATTAATGTCCACAGGTGAAAGGATTTCTCTTCG 1167
 Db 181 gatacagggggcagggagggagcgtcgtcttagtcttgaagaggtcagacacagaaga 240
 Cp 1166 GATAGCAGGGGAGGAGGAGAGCGTCGCTTAGTCTTTGAAGAGGTCAGACACAGAA 1107
 Db 241 aactatgcagagggccaccaggggctccttgagcaaatcctgccagaacatcactggatg 300
 Cp 1106 AACATATCAGCAGGCCACACAGGGCTCTTGTAGCAAAATCCTGCCAGAACATCCTGGGATG 1047
 Db 301 gtgcttgcgtgatacgcgagacagtcctcgtgtagaaggccatcatgatacgaagtga 360
 Cp 1046 GTGCTTGTGTCTGATACGCGAGACAGTCCCTGTGTAGAAGGCCATCATGATC-AAGGTGA 988
 Db 361 actccaggaagggggcggagcggggtcctcctcgcgaagtnaagcgccctgcaggt 420
 Cp 987 ACTGC-AGGA-GGGGCCGGAGCAGCGGGGTCTCT-CCCAAGTGAAGCGGCCCTGCAGGT 931
 Db 421 atagcaccataacagcatagtgatgaggaagagagggcagggagagaagaagga 480
 Cp 930 ATAGCACCACCAATACAGCATAGTGATGATGGA-GAAGGAGGCATGCCAG-AGAAGAAGGA 873
 Db 481 ctctcgggttccctggaatttgcgtggttcacacotcgtcactctgttagtctgaaatg 540
 Cp 872 CTCTCTGG-CTTCC-TGGACTTT-GCTG-TCATCACCTCTGCATCTGTAGTTCTGAA-TG 818
 Db 541 taaccttcagagcagttgattcgg 565
 Cp 817 TAGCC-TTCAGAGCAGTTGATCTGG 794

RESULT 4
 LOCUS W70040 449 bp mRNA EST 16-OCT-1996
 DEFINITION z449b06.s1 Soares fetal heart NBHH19W Homo sapiens cDNA clone 343955 3'.
 ACCESSION W70040
 NID g1379321
 KEYWORDS EST.

```

SOURCE
ORGANISM      human.
Eukaryotae:  mitochondrial eukaryotes: Metazoa: Chordata:
Vertebrata:  Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 449)
REFERENCE
AUTHORS      Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Wohldmann,P. and
Wilson,R.
TITLE        The WashU-Merck EST Project
JOURNAL      Unpublished (1995)
COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1875 Std Error: 0.00
Seq primer: mob.REGA-ET
High quality sequence stop: 227.
Location/Qualifiers
1..449
/organism="Homo sapiens"
/note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCATCTTTTGTGTTTGTGTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NHLL19W."
/clone="343955"
/clone_lib="Soares fetal heart NBHH19W"
/dev_stage="19 weeks"
/sex="unknown"
/lab_host="DH10B (ampicillin resistant)"
complement(<1..>449) 98 t

mRNA
BASE COUNT  123 a  98 c  130 g  98 t
ORIGIN

Query Match      32.3%; Score 440; DB 173; Length 449;
Best Local Similarity 99.6%; Pred. No. 0.00e+00;
Matches 447; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Db 1  taaaaatcagtcgggcaaaagtgtttccctacattctactgtctgatgagattggagagc 60
Cp 1347  TAAAAATCAGTCGGGCAAAAGTGTTCCTACATCTCTACTGTCTGTGATGAGATTGGAGAGC 1288
Db 61  agcaagaactgtgtcagcagtcatttttcaaaaaaacagctcaggagtggtggcgacct 120
Cp 1287  AGCAAGAACTGTGTCTCAGCAGTCATTTTACAAAAACAGCTCAGGAGGTGGGTGGCACCT 1228
Db 121  acatcatgttggtagttgtctgtcgaataatagcccacagtgtaaaagatttccttc 180
Cp 1227  ACATCATGTTGGTGTGATTGTTCTGTCATATAT-GTCCACAGGTGAAGAGATTTCCTTC 1169
Db 181  cggatagcaggggcagggcagggagcgtgtcttagtcttgaagaggtcagacacgaag 240
Cp 1168  CGGATAGCAGGGGCAGGCAGGAGCGTCTGCTTAGTCTTTGAAGAGGTGAGAGTTCAGACGAG 1109
Db 241  aaaaactatgcagcagggccacagggtcccttgagcaaatcctccgaaacatcactcggga 300
Cp 1108  AAAAATATGAGCAGGCCACAGGGCTCTTTGAGCAAAATCTGCCAAGATCACTGGGA 1049

```

```

Db 301  tgggtgcttggtctgtacacgcagacagtcctcgtgtataagcccatcatgatcaggtg 360
Cp 1048  TGGTGCTTGTGCTGTATACGGGACAGCAGTCCTCGGTGTAGAACCCCATCATGATCAAGGTG 989
Db 361  aactgcaggagggccggcagcaggcggtcctctcccaagtgaagcggcctgcaggtat 420
Cp 988  AACTGCAGGAGGGCGGAGCAGCGGGCTCTCTGCCAAGTGAAGCGGCGCTGCAGGTAT 929
Db 421  agcaccacaatacagcatagtgtacatgga 449
Cp 928  AGCACCACAATACAGCATAGTGTACATGGA 900
RESULT 5
LOCUS      H50533              451 bp      mRNA      EST      18-SEP-1995
DEFINITION ypo8a11.s1 Homo sapiens cDNA clone 186812 3'.
ACCESSION  H50533
NID        g990374
KEYWORDS   EST.
SOURCE     human clone-186812 primer=Promega -21ml3 library=Soares breast
3NHBST vector-pT7T3D (Pharmacia) with a modified polylinker
host=DH10B (ampicillin resistant) Rsite1=Not I Rsite2=Eco RI Adult
human. 1st strand cDNA was primed with a Not I - oligo(dT) primer
[5' TGTTACCAATCTGAAGTGGAGCGCGCATCTTTTGTGTTTGTGTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia),
digested with Not I and cloned into the Not I and Eco RI sites of a
modified pT7T3 vector (Pharmacia). Library went through one round
of normalization to a Cot = 20. Library constructed by Bento Soares
and M.Fatima Bonaldo.
ORGANISM   Homo sapiens
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE  1 (bases 1 to 451)
AUTHORS    Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Wohldmann,P. and
Wilson,R.
TITLE      The WashU-Merck EST Project
JOURNAL    Unpublished (1995)
COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 273
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES
Source
1..451
/organism="Homo sapiens"
/clone="186812"
<1..>451
mRNA
BASE COUNT  120 a  98 c  135 g  94 t  4 others
ORIGIN

Query Match      29.7%; Score 405; DB 83; Length 451;
Best Local Similarity 97.8%; Pred. No. 0.00e+00;
Matches 442; Conservative 0; Mismatches 4; Indels 6; Gaps 6;

Db 1  ttccctacattctactgtctgatgattggagagcagcaagaacttctgttcagcag 60
Cp 1324  TTCCCTACATTCCTACTGCTGTCTGATGAGATTGGAGACACAAAGACTTCTGT-CAGCAG 1266
Db 61  ttcatttcaaaaaacagctcaggaggtgggtgggncaccctacatcatgttgggtgattg 120
Cp 1265  T-CATTTTACAAAAACAGCTCAGGAGGTGGTGG-CACCTACATCATGTTGTGGTGATTG 1208

```

```

Db 121 ttctgttcataatgtccacaggtgaagaggatttcttcccgatagcaggggagggca 180
Cp 1207 TTCCTGTCAATAATGTCCACAGGTGAAAGGATTTCTTCCGGATAG-CAGGGCGAGG-CA 1150

Db 181 gggagagcgtcttagtcttgaagaggtcagacagagaagaactatgacagagccca 240
Cp 1149 GGGAGAGCGTCTTGTAGCTTTGAAGAGGTCAGACAGCAAGAAACTATGACAGAGGCCA 1090

Db 241 ccagggctcttgagcaaatcctgcagaacatcactggggtgcttgggctgata 300
Cp 1089 CCAGGCTCTTGAGCAATCTGCCAGAACATCAGTGGGATGTGCTTGTGCTGATA 1030

Db 301 cgcgagacagtcctcgttagaagggccatcatgatcaaggtgaaactnagggggccgga 360
Cp 1029 CGCGAGACAGTCCCGTGTGAGAGGCGCATCATGATCAAGGTGAAGTGCAGGAGGGCCGGA 970

Db 361 gcagcgggctcttcgcaagtcgaagtcgagggcctnaggatagaccacaaatacagcatag 420
Cp 969 CGAGCGGGCTCTTCGCCAAGTGAAGCGGGCTGCAGGTATAGCACCAATATACAGCATAG 910

Db 421 tgtacatgagaagaggagcagtg-ccagsgaag 451
Cp 909 TGTACATGAGAGAGGAGCATGGCCAGAGAG 878

RESULT 6
LOCUS H04639 480 bp mRNA EST 20-JUN-1995
DEFINITION Y149a08.s1 Homo sapiens cDNA clone 152054 3'.
ACCESSION H04639
NID 9867572
KEYWORDS EST.
SOURCE human clone=152054 library=Soares placenta Nb2HP vector-ptT73D
(Pharmacia) with a modified polylinker host=DH10B (ampicillin
resistant) primer=Pronega -21ml3 Rsite1=Not I Rsite2=Eco RI Female
placenta obtained at birth (full term). 1st strand cDNA was primed
with a Not I - oligo(dT) primer [5'
AACTGGAAGAAATCGCGCGCGAGGAATTTTTTTTTTTT 3'], double-stranded
cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the modified pT73
vector. library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo.
Homo sapiens
ORGANISM
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 480)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 270
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES
source Location/Qualifiers
1..480
/organism="Homo sapiens"
/clone="152054"
121 a 107 c 133 g 112 t 7 others
BASE COUNT 121 a 107 c 133 g 112 t 7 others
ORIGIN

```

```

Query Match 29.5%; Score 402; DB 54; Length 480;
Best Local Similarity 97.7%; Pred. No. 0.00e+00;
Matches 422; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

Db 1 ttctctacattctactgtctgtatgattggagagcagcaagaacttgtgtcagcagt 60
Cp 1324 TTTCCCTACATTTCTACTGTCTGTATGATTGGAGAGCAGCAAGAACTTGTCTGACGAGT 1265

Db 61 catttacaacacagctcagaggggtgggtggcaccatcacatcatgttgtgtatgttc 120
Cp 1264 CATTTTCAAAAACAGCTCAGGAGGTGGGTGGCCACTTACATCATGTGTGTGATTGTTT 1205

Db 121 ctgtcaataatgtccacaggtgaaagatttcttctccgtagcaggggagcagggag 180
Cp 1204 CTGTCAATAATGTCCACAGGTGAAAGGATTTCTTCCGGATAGCAGGGCGAGGAG 1145

Db 181 aggtctcttagtcttgaagaggtcagacacagaagaaactatgcagagggccaccagg 240
Cp 1144 AGCGTCTCTTAGTCTTTGAAGAGGTCTAGACACGAAGAAAACTATGCAGCAGGCCACGAG 1085

Db 241 gctccttgagcaaatcctgcagaacatcactggatgggtgcttntgtctgatacgcga 300
Cp 1084 GTCCTTGTAGCAAAATCTCTCCAGAACATCACTGGGATGGTCTGTGTCTGTATACCCGA 1025

Db 301 gacagtcctcgttagaaggccatcatcatcaagtgtaactgcagggagggccagncagg 360
Cp 1024 GACAGTCCCGTGTAGAGGCCCATCATGATCAAGGTGAACCTGCAGGAGGGCCGAGCAGG 965

Db 361 cnggctcct-gcaagtgagcgtgcctnccaggtat-gcaccacaaatacagcatagt-tac 417
Cp 964 CGGCTCTCTCGCCAAAGTGAAGCGGGCTGCAGGTATATAGCACCACCAATACAGCATAGTGTAC 905

Db 418 atggagnagag 429
Cp 904 ATGGAGGAAGGAG 893

RESULT 7
LOCUS AA156340 428 bp mRNA EST 11-DEC-1996
DEFINITION 2054f04.s1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
590719 3'.
ACCESSION AA156340
NID gi727974
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 428)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 329.
FEATURES
source Location/Qualifiers
1..428
/organism="Homo sapiens"
/note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
ECORI; Site_2: XhoI; Cloned unidirectionally. Primer:

```

Oligo dT. Pancreatic adenocarcinoma cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3' "

/clone="590719"
/lab_host="SOLR cells (kanamycin resistant)"
complement(<1...>428)

MRNA
BASE COUNT 113 a 95 c 125 g 93 t 2 others
ORIGIN

Query Match 29.4%; Score 401; DB 191; Length 428;
Best Local Similarity 98.8%; Pred. No. 0.00e+00;
Matches 419; Conservative 0; Mismatches 2; Indels 3; Gaps 3;

Db 1 aaaaatcagtcggcgaagtttttccctacattctactgtctgatgagattgagagcag 60
Cp 1345 AAAATCAGTCGGGCAAAAGTTTTCCTACATCTACTGTCTGATGAGATTGGAGAGCAG 1286
Db 61 caagaacttgctgcagcagtcattttacaaaacagctcagaggtggtggcactac 120
Cp 1285 CAAGAACTTCTGTCAGCAGTCATTTTACAAAACAGCTCAGGAGGTGGGTGCACCTAC 1226
Db 121 atcatgttggtgattgttcctctcgaataatgtccacaggtgaaaggatttctctccgg 180
Cp 1225 ATCATGTTGTGGTGTGTTCTCTGCAATATGTCACAGGTGAAAGGATTCTCTCCGG 1166
Db 181 atagcagggcagcagcagcagcgtctttagtttgaagggtcagacagaagaaa 240
Cp 1165 ATAGCAGGGCAGCGAGGAGAGCGTCGTCTTAGTCTTGAAGAGGTCAGACAGAA 1106
Db 241 actatgcagcagggccacaggggtctcttgagcaaatctcccgaaacatcactgggatgg 300
Cp 1105 ACTATGCAGAGCCACCAGGGTCTCTTGAGCAATCTCTCCAGACATCTGGGATGG 1046
Db 301 tggcttggtcctgatacgcagacagtcctcgtgtagaagggccatcatgatcaagtgaa 360
Cp 1045 TGCTTGTGTC-TGATACGGGAGACAGTCCTCGTGTAGAAGGCCATCATGATCAAGGTGA 987
Db 361 ctacagagggccgagcagcagcgggtctctccgaattgaaagcgggnccttcaggtat 420
Cp 986 CTCAGAGGGGCGGAGCAGCGGGGCTCTCTGCCAAGT-GAAGCGGGCCT-GCAGGTAT 929
Db 421 agca 424
Cp 928 AGCA 925

RESULT 8
LOCUS N68923 413 bp mRNA EST 13-MAR-1996
DEFINITION za69f12.s1 Homo sapiens cDNA clone 297839 3'.
ACCESSION N68923
NID 91225084
KEYWORDS EST.
SOURCE human clone=297839 primer=ml3 -40 forward library=Soares fetal lung

NbH19W vector=PT73D (Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) Rsite1=Not I Rsite2=Eco RI 19 week fetus. 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTACCAATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo.

Homo sapiens

ORGANISM

REFERENCE 1 (bases 1 to 413)

AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlffing,T., Soares,M., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
The WashU-Merck EST Project

TITLE

JOURNAL COMMENT

Unpublished (1995)

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Location/Qualifiers

1..413

/organism="Homo sapiens"

/clone="297839"

<1...>413

BASE COUNT 107 a 92 c 124 g 90 t

ORIGIN

Query Match 28.5%; Score 388; DB 102; Length 413;
Best Local Similarity 99.3%; Pred. No. 0.00e+00;
Matches 406; Conservative 0; Mismatches 0; Indels 3; Gaps 3;

Db 1 aaaaatcagtcggcgaagtttttccctacattctactgtctgatgagattgagagcag 60
Cp 1345 AAAATCAGTCGGGCAAAAGTTTTCCTACATCTACTGTCTGATGAGATTGGAGAGCAG 1286
Db 61 caagaacttgctgcagcagtcattttacaaaacagctcagaggtggtggcactac 120
Cp 1285 CAAGAACTTCTGTCAGCAGTCATTTTACAAAACAGCTCAGGAGGTGGGTGCACCTAC 1226
Db 121 atcatgttggtgattgttcctctcgaataatgtccacaggtgaaaggatttctctccgg 180
Cp 1225 ATCATGTTGTGGTGTGTTCTCTGCAATATGTCACAGGTGAAAGGATTCTCTCCGG 1166
Db 181 atagcagggcagcagcagcagcgtctttagtttgaagggtcagacagaagaaa 240
Cp 1165 ATAGCAGGGCAGCGAGGAGAGCGTCGTCTTAGTCTTGAAGAGGTCAGACAGAA 1106
Db 241 atatgcagcagggccacaggggtctcttgagcaaatctcccgaaacatcactgggatg 300
Cp 1105 ACTATGCAGAGCCACCAGGGTCTCTTGAGCAATCTCTCCAGACATCTGGGATG 1047
Db 301 tggcttggtcctgatacgcagacagtcctcgtgtagaagggccatcatgatcaagtgaa 360
Cp 1046 TGCTTGTGTC-TGATACGGGAGACAGTCCTCGTGTAGAAGGCCATCATG-ATCAAGGT-G 989
Db 361 aactgcagagggcggcggagcagcggggtctctccgaattgaaagcggg 409
Cp 988 AACTGCAGAGGGCGGAGAGGAGCGGGGCTCTCTGCCAAGTGAAGCGGG 940

RESULT 9
LOCUS N70674 402 bp mRNA EST 14-MAR-1996
DEFINITION za32a04.s1 Homo sapiens cDNA clone 294222 3'.
ACCESSION N70674
NID 91227254
KEYWORDS EST.
SOURCE human clone=294222 primer=ml3 -40 forward library=Soares fetal

liver spleen INFUS vector=PT73D (Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) Rsite1=Pac I Rsite2=Eco RI Liver and spleen from a 20 week post conception male fetus. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5'-AACTGGGAATAATTAAGATCTTTTCTTTTCTTTT-3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

Homo sapiens

ORGANISM

Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;

Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
AUTHORS

TITLE	COMMENT
WILSON, R. The WashU-Merck EST Project Unpublished (1995)	

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Tel: 314 286 1800
Fax: 314 286 1810

FEATURES
SOURCE

```

source
1..402
/organism="Homo sapiens"
/clone="294222"
<1 >402
mRNA

```

BASE COUNT		ORIGIN	
106 a	88 c	109 g	91 t
8 others		8 others	

Query Match 27.9%; Score 380; DB 103; Length 402;
Best Local Similarity 97.5%; Pred. No. 0.00e+00;
Matches 392; Conservative 0; Mismatches 8; Indels 2; Gaps 2;

Db 1 ttaaaatcagtcgggcaaaagttttccctacattctactgtctgatgagattggagag 60

1348 TTAATAATCAGTCGGGCGAAAAGTTTTTCCCTACATTCTACTGTCTGATGAGATTGGAGAG 1289

Db 61 cagcaagaacttctcagcagtcattttacaaaaacagctcangaggtgggtggcacc 120

1288 CAGCAAGAACCTTGCTGTCAGCAGTCATTTTACAAAAACAGCTCAGGAGGTGGGTGGCACC 1229

Db 121 tacatcatgttggtgattgttcctgtcaataatntccacaggtgaaggatttccttc 180

1228 TACATCATGTTGTGGTGATTTGTTCCCTGTCAATAATGTCCACAGGTGAAGGATTTCCCTTC

Db 181 cggatagcaggggcaggcaggagagcgtcgtcttagtcttgaanaggtcagacacgaag 240

CP 1168 CGGATAGCAGGGGCAGCGCAGGGAGCGTCGCTTAGTCTTGAAGAGGTCAGACACGAAG 1109

Db 241 aaactatgcagcaggccaccagggtccttgagcaaatcctgccagaacatcactngga 300

Cp 1108 AAAACTATGCAGCAGGCCACAGGGCTCCTTGAGCAAATCCTGCCAGACATCACATGGGA 1049

DG 301 cggcgcttgcgatacgcganacagtcgcgintagaaggccatcatgatcaaggt 360

Cp 1048 TGGTGC TTGT -GGTCTGATACGGCAGACAGTCCCGTGTAGAAGGCCATCATGATCAAGGT 990

[illegible]

cp 989 GAACTGCAGGAGGGGCCGGAGCAGGC -GGGCTCCTCGCCAAG 949

RESULT	10
LOCUS	R63796 446 bp mRNA EST 26-MAY-1995
DEFINITION	Y115H04.r1 Homo sapiens cDNA clone 139351 5'.
ACCESSION	R63796
NID	g835675
KEYWORDS	EST.
SOURCE	human clone-139351 library=Soares placenta NB2HP vector=pT73D (Pharmacia) with a modified polylinker host-DHI0B (ampicillin resistant), primer=M13Rp1 Rsicle2-Not I Rsiite2-Eco RI Female placenta obtained at birth (full term). 1st strand cDNA was primed with a

```

RESULT 11
LOCUS      R63741      431 bp      mRNA      EST      26-MAY-1995
DEFINITION Y115H04.s1 Homo sapiens cDNA clone 139351 3'.
ACCESSION  R63741
NID        Q835620
KEYWORDS   EST.
SOURCE     human clone-139351 library=Soares placenta Nb2HP vector=pt7T3D
            (Pharmacia) with a modified polylinker host=DH10B (ampicillin
            resistant) primer=Promega -21m13 Rsite1-Not I Rsite2-Eco RI Female
            placenta obtained at birth (full term). 1st strand cDNA was primed
            with a Not I - oligo(GT) primer [5',
            AACTCGGAAGTAATTCGGCGCGAGGAATTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not
            I and Cloned into the Not I and Eco RI sites of the modified pT7T3
            vector. Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
ORGANISM   Homo sapiens
            Eukaryote; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Chonadata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1. (bases 1 to 431)
            Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevasaki,E., Waterston,R., Williamson,A., Wohlmann,P. and
            Wilson,R.
            The WashU-Merck EST Project
            Unpublished (1995)
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@wustl.wustl.edu
            High quality sequence stops: 100
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES   source
            Location/Qualifiers
            1..431
            /organism="Homo sapiens"
            /clone="139351"
BASE COUNT 111 a 91 C 123 G 97 T 9 others
ORIGIN
Query Match 26.2%; Score 357; DB 29; Length 431;
Best Local Similarity 96.0%; Pred. No. 0.00e+00;
Matches 406; Conservative 0; Mismatches 9; Indels 8; Gaps 8;
Db 1 aaatcagtcggcgaaagattttccctacattctactgtctgatgagatggagacgag 60
Cp 1345 AATATCATGTCGGGCAAGAATTTTCCCTACATTTCTACTGTCTGATGAGATGGAGAGC 1286
Db 61 caagaacttgctgcagcagtcattttacaaaaacagctcagagggtgggtggcacctac 120
Cp 1385 CAAGAACTTGCTGTCAGCAGTCATTTACAAAAACAGCTCAGGAGGTGGGTGGCACCTAC 1226
Db 121 atcatgttggtgatgttctcgtcgaataatgtccacaggtgaaaggatttccttcgg 180
Cp 1225 ATCATGTTGTGGTGATTGTTCCTGTCATAATATGCCACAGGTGAAGGATTTCCTCCGG 1166
Db 181 atagcaggggcagcgaggagcgcgtgcttagctcttgagagaggtcagacacaaagaa 240
Cp 1165 ATACAGGGGCGAGCAGGAGGAGGCTGCTCTTAGTCTTCAAGAGGTCAGACACGAAGAAA 1106
Db 241 actatgcagcagggccacacagggctcctcttgagcaaatcttcgcagaacatcactggatgg 300
Cp 1105 ACTATGACAGAGGCCACAGGGCTCCCTTGACAAATCTGTCAGAAACATCACTGGGATGG 1046

```

Db	301	gtgcttgagggtctnatacgcgagacagtcctccgtntagggaagccattcatatgatacaggt	360
Cp	1045	-TGTCTTGG-TCTGTATACGCGAGACATCCCGTGTAG-AAGGCCAT-CATGATCAAGGT	990
Db	361	gaacttcaggaggccngagcagcngcctncttcgccaagttnaagaggccctnca	420
Cp	989	GAACT-CGAGG-AGGGCCGAGCAGCGCGGCTCT-CGCCAAGT-GAAGCGGCCCTGCA	934
Db	421	ggt 423	
Cp	933	GGT 931	
RESULT	12		
LOCUS	H02469	463 bp	EST
DEFINITION	Y135409.s1 Homo sapiens cDNA clone 150737 3'.		20-JUN-1995
ACCESSION	H02469		
NID	9865402		
KEYWORDS	EST.		
SOURCE	human clone=150737 library=Soares placenta ND2HP vector=pt7T3D (Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) primer=Promega -31m3 Reitel=Not I Raites=Eco RI Female placenta obtained at birth (full term). 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAATTCGGCCGCGAGGAATTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomes; Sarcoteriygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homidae; Homo.		
AUTHORS	1 (bases 1 to 463) Hallier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.		
TITLE	The Washu-Merck EST Project		
JOURNAL	Unpublished (1995)		
COMMENT	Contact: Wilson RK Washu-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu High quality sequence stops: 358 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. IMAGE Consortium (info@image.llnl.gov) for further information. Location/Qualifiers 1..463 /organism="Homo sapiens" /clone="150737"		
FEATURES	source		
BASE COUNT	122 a 96 c 134 g 103 t	8 others	
ORIGIN			
Query Match	25.9%;	Score 353;	DB 43; Length 463;
Best Local Similarity	94.9%;	Pred. NO. 0.00e+00;	
Matches	429; Conservative	0; Mismatches 11; Indels 12; Gaps 11;	
Db	1	aaatcagttgttaaaggttttccctacattctactctgtctgtatgagttgagagcag	60
Cp	1345	AAATCATCTCGGGCAAAAGTTTTTCCCTACATTCTACTGCTGATGAGATTGGAGACG	1286
Db	61	caagaacttgctgcagcagtcatttcacaaaaaacagctcaggagggtgggtggcacctac	120
Cp	1285	CAGAAGCTTGCTGTCAGCAGTCATTTTACAAAACAGCTCAGGAGGTGGTGACCTAC	1226

Db 121 atcatgttggtgatttcttctgtcaataatgtccacagtgtaaaagatttcttcgg 180
 |||||
 Cp 1225 ATCATGTTGTTGTTGTTCTCTGTCATAATGTCACAGGTGAAGGATTCTCTCCGG 1166
 |||||
 Db 181 atagcaggggagcagggagagcgtctcttagtctctgaagaggtcagacacaga 240
 |||||
 Cp 1165 ATAGCAGGGGAGCGAGGAGAGCGTCTCTTAGTCTTTGAAGAGGTCAGACACGA 1106
 |||||
 Db 241 actatgcagcagggccacagggctcttgagcaaatctctgccaaacatcactgggatg 300
 |||||
 Cp 1105 ACTATGCAGCAGGCCACAGGGCTCTTGTAGCAATCTCTGCCAGAACATCACTGGG-ATG 1047
 |||||
 Db 301 ggtgcttggtgtgattacagcgagacagtcctccgtgtaggaagccattcatgacaaa 360
 |||||
 Cp 1046 G-TGCTTCTGG-TCGTATACG-CGAGACAGTCCCCTGTAG-AAGGCCAT-CATGATCAA- 993
 |||||
 Db 361 ggggtgaactgcagggagggcngagcagcggtctctctnccaaagtgagcgggcctn 420
 |||||
 Cp 992 -GGTGAACGTGCAGGAGGGGGCGGAGCAGCGGGCT-CCTGCCCAAGT-GAAGCGGGCCTG 936
 |||||
 Db 421 ccaggttttaggcacnaatcacgactnata 452
 |||||
 Cp 935 C-AGGTATAG-CACCAATAACAGCATGTGTA 906
 |||||

RESULT 13
 LOCUS H04659 348 bp mRNA EST 20-JUN-1995
 DEFINITION Y149a08.r1 Homo sapiens CDNA clone 152054 5'.
 ACCESSION H04659
 NID 9867592
 KEYWORDS EST.
 SOURCE human clone=152054 library=Soares placenta Nb2HP vector-pt7T3D (Pharmacia) with a modified polylinker host-DH10B (ampicillin resistant) primer-W13Rev Rsite1-Not I Rsite2-Eco RI Female placenta obtained at birth (full term). 1st strand CDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAAGATTCGCGCGCGAGGAATTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM Homo sapiens
 Eukaryotes; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 348)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 High quality sequence stops: 257
 Source: IMAGE Consortium, LLNL
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Location/Qualifiers
 1..348
 /organism="Homo sapiens"
 /clone="152054"

BASE COUNT 74 a 102 c 79 g 90 t 3 others
 ORIGIN

FEATURES
 SOURCE

Query Match 24.0%; Score 327; DB 54; Length 348;
 Best Local Similarity 98.0%; Pred. No. 0.00e+00;
 Matches 341; Conservative 0; Mismatches 5; Indels 2; Gaps 2;
 Db 1 gccagataactgtctgaaggtacattcagaactacagatgcagaggatgacanca 60
 QY 793 GCCAGATCAACTGCTCTGAAGGCTACATTTCAGAACTACAGATGCAGAGGTGATGACAGCA 852
 |||||
 Db 61 aagtcaggaagccaggaagtccttctcttgccatgcctctcttccatgacacta 120
 |||||
 QY 853 AAGTCCAGGAAGCCAGGAAGTCTTCTCTTGCCATGCTCTCTTCCATGTACACTA 912
 |||||
 Db 121 tgcgtatttngtctatccctcagcccgcttcacctttgagggagggccgcctgtccc 180
 |||||
 QY 913 TGCCTGATTGTTGCTACTACTCTGCAGCCCGCTTCACTTGGCGAGGAGCCGCTGCTCC 972
 |||||
 Db 181 ngccctctcgcagttcacctctgatgatggcctcttaacacgggactgtctgcgtat 240
 |||||
 QY 973 GGGCCCTCTGTCAGTTTCACTTGTATGATGATGGCTTCTACACGGGACTGTCTCGCGTAT 1032
 |||||
 Db 241 cagaccacagcaccatccacagtcagttcttctgagggatttgcacaggagccctggtg 300
 |||||
 QY 1033 CAGACCACAGCACCATCCCAAGTGTCTTGGCAGGATTTGCTCAAGGAGCCCTGTGTG 1092
 |||||
 Db 301 cctgctgcatagttttctt-gtgtctgacct-ttcaagactaaagcga 346
 |||||
 QY 1093 CCTGCTGCATAGTTTCTTCTGCTGCTGACCTTTCAGACTTAAGACGA 1140
 |||||

RESULT 14
 LOCUS H02468 475 bp mRNA EST 20-JUN-1995
 DEFINITION Y135d09.r1 Homo sapiens CDNA clone 150737 5'.
 ACCESSION H02468
 NID 9865401
 KEYWORDS EST.
 SOURCE human clone=150737 library=Soares placenta Nb2HP vector-pt7T3D (Pharmacia) with a modified polylinker host-DH10B (ampicillin resistant) primer-W13Rpl Rsite1-Not I Rsite2-Eco RI Female placenta obtained at birth (full term). 1st strand CDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAAGATTCGCGCGCGAGGAATTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM Homo sapiens
 Eukaryotes; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 475)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 High quality sequence stops: 321
 Source: IMAGE Consortium, LLNL
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Location/Qualifiers
 1..475

FEATURES
 SOURCE

W P S R L A

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by Intelligent, Inc.

MPSrch_on n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Thu Nov 6 12:15:43 1997; Maspar time 370.56 Seconds
Tabular output not generated. 1010.069 Million cell updates/sec

Title: >US-08-842-827-5
Description: (1-1362) from US08842827.seq
Perfect Score: 1362
N.A. Sequence: 1 GCGCGAGCTCGCAAAAGTT.....TTTTAAAAAATAAAAAA 1362
Comp: CCGCGTCGAGCGTTTCA.....AAATTTTTTTTTTTTT

Scoring table: TABLE default
Gap 6
Nmatch STD : Dbase 0; Query 0
Searched: 359085 seqs, 137405154 bases x 2
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: EST-STS-THREE
1:EST199 2:EST200 3:EST201 4:EST202 5:EST203 6:EST204
7:EST205 8:EST206 9:EST207 10:EST208 11:EST209 12:EST210
13:EST211 14:EST212 15:EST213 16:EST214 17:EST215
18:EST216 19:EST217 20:EST218 21:EST219 22:EST220
23:EST221 24:EST222 25:EST223 26:EST224 27:EST225
28:EST226 29:EST227 30:EST228 31:EST229 32:EST230
33:EST231 34:EST232 35:EST233 36:STS1 37:STS2 38:STS3
39:STS4 40:STS5 41:STS6 42:STS7 43:STS8 44:STS9 45:STS10
46:STS11 47:STS12 48:STS13
EST-STS-FOUR
49:gnEST1 50:gnEST2 51:gnEST3 52:gnEST4 53:gnEST5
54:gnEST6 55:gnEST7 56:gnEST8 57:gnEST9 58:gnEST10
59:gnEST11 60:gnEST12 61:gnEST13 62:gnEST14 63:gnEST15
64:gnEST16 65:gnEST17 66:gnEST18 67:gnEST19 68:gnEST20
69:gnEST21 70:gnEST22 71:gnEST23 72:gnEST24 73:gnEST25
74:gnEST26 75:gnEST27 76:gnEST28 77:gnEST29 78:gnEST30
79:gnEST31 80:gnEST32 81:gnEST33 82:gnEST34 83:gnEST35
84:gnEST36 85:gnEST37 86:gnEST38 87:gnEST39 88:gnEST40
89:gnEST41 90:gnEST42 91:gnEST43 92:gnEST44 93:gnEST45
94:gnEST46 95:gnEST47 96:gnEST48 97:gnEST49 98:gnEST50
99:gnEST51 100:gnEST52 101:gnEST53 102:gnEST54
103:gnEST55 104:gnEST56 105:gnEST57 106:gnEST58 107:gnEST59
108:gnEST60

Database: EST-STS-FOUR
49:gnEST1 50:gnEST2 51:gnEST3 52:gnEST4 53:gnEST5
54:gnEST6 55:gnEST7 56:gnEST8 57:gnEST9 58:gnEST10
59:gnEST11 60:gnEST12 61:gnEST13 62:gnEST14 63:gnEST15
64:gnEST16 65:gnEST17 66:gnEST18 67:gnEST19 68:gnEST20
69:gnEST21 70:gnEST22 71:gnEST23 72:gnEST24 73:gnEST25
74:gnEST26 75:gnEST27 76:gnEST28 77:gnEST29 78:gnEST30
79:gnEST31 80:gnEST32 81:gnEST33 82:gnEST34 83:gnEST35
84:gnEST36 85:gnEST37 86:gnEST38 87:gnEST39 88:gnEST40
89:gnEST41 90:gnEST42 91:gnEST43 92:gnEST44 93:gnEST45
94:gnEST46 95:gnEST47 96:gnEST48 97:gnEST49 98:gnEST50
99:gnEST51 100:gnEST52 101:gnEST53 102:gnEST54
103:gnEST55 104:gnEST56 105:gnEST57 106:gnEST58 107:gnEST59
108:gnEST60

Statistics: Mean 11.582; Variance 2.791; scale 4.149
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result Query
SUMMARIES

No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	500	36.7	564	1	W30942	zc64f09.r1 Soares fet	0.00e+00
2	500	36.7	564	88	HS942332	zc64f09.r1 Soares fet	0.00e+00
3	382	28.0	441	15	AA192756	zq12c05.s1 Stratagene	0.00e+00
4	376	27.6	403	87	HS733324	zc64f09.s1 Soares fet	0.00e+00
5	376	27.6	403	1	W02733	zc64f09.s1 Soares fet	0.00e+00
6	336	24.6	446	12	AA181403	zp52f10.s1 Stratagene	0.00e+00
7	250	18.4	402	27	W39811	307 Mouse VM cDNA lib	0.00e+00
8	244	17.9	437	27	W39816	309 Mouse VM cDNA lib	0.00e+00
9	243	17.8	546	40	GI3351	human STS WI-12145	0.00e+00
10	225	16.5	421	8	AA126964	zl87e03.s1 Stratagene	0.00e+00
11	211	15.5	374	1	W23530	zc71h10.s1 Soares fet	4.06e-296
12	211	15.5	374	87	HS530338	zc71h10.s1 Soares fet	4.06e-296
13	190	14.0	242	27	W39815	308 Mouse VM cDNA lib	2.04e-260
14	179	13.1	236	81	HS1146241	zp52f10.r1 Stratagene	8.22e-242
15	179	13.1	236	12	AA182434	zp52f10.r1 Stratagene	8.22e-242
16	133	9.8	440	55	AA105463	mn92b06.r1 Stratagene	5.56e-165
17	124	9.1	467	10	AA016796	mh43e01.r1 Soares mou	3.42e-120
18	110	8.1	413	21	AA021853	mh86b01.r1 Soares mou	2.08e-127
19	78	5.7	385	33	AA107326	ml95b10.r1 Stratagene	9.47e-77
20	74	5.4	437	33	AA106725	mm17c02.r1 Stratagene	1.30e-70
21	46	3.4	478	26	N75714	yv31b01.r1 Soares fet	7.98e-30
22	45	3.3	914	34	AA141702	CK02248.contig Drosop	1.81e-28
23	39	2.9	422	34	AA142082	CK00216.3prime Drosop	1.47e-20
24	34	2.5	87	33	AA106892	ml85a07.r1 Stratagene	2.51e-14
25	26	1.9	569	75	AT0465	7309 Arabidopsis thal	2.53e-05
26	24	1.8	361	4	AA115689	zk96f04.s1 Soares pre	2.51e-03
27	24	1.8	533	19	AA206113	zq32g06.r1 Stratagene	2.51e-03
28	24	1.8	593	75	AT0302	4038 Arabidopsis thal	2.51e-03
29	23	1.7	73	48	HUMUT934A	Human STS UT934, 5' p	2.22e-02
30	23	1.7	157	71	AA251615	zs08h10.r1 Soares Nhh	2.22e-02
31	23	1.7	296	72	AA254840	mz75c09.r1 Soares mou	2.22e-02
32	23	1.7	296	93	MM1171194	mz75c09.r1 Soares mou	2.22e-02
33	23	1.7	345	99	MAA31464	mw38e02.r1 Soares mou	2.22e-02
34	23	1.7	345	64	AA231464	mw38e02.r1 Soares mou	2.22e-02
35	23	1.7	357	83	HS1151833	zr47e08.s1 Soares Nhh	2.22e-02
36	23	1.7	371	105	TB0691	T2504 MVAR4 bloodstre	2.22e-02
37	23	1.7	403	93	MM1164661	mz58f02.r1 Soares mou	2.22e-02
38	23	1.7	431	14	AA188740	zp78d02.r1 Stratagene	2.22e-02
39	23	1.7	434	84	HS1160099	zr67d03.s1 Soares Nhh	2.22e-02
40	23	1.7	455	87	HS577332	zb92b11.s1 Soares par	2.22e-02
41	23	1.7	475	30	AA061089	ml43h09.r1 Stratagene	2.22e-02
42	23	1.7	488	6	AA121386	zn77f06.s1 Stratagene	2.22e-02
43	23	1.7	490	9	AA133331	zn92e01.s1 Stratagene	2.22e-02
44	23	1.7	519	22	AA213648	mu80c03.r1 Stratagene	2.22e-02
45	23	1.7	528	53	AA193447	zr40e04.r1 Soares Nhh	2.22e-02

ALIGNMENTS

RESULT 1
LOCUS W30942 564 bp mRNA EST 25-NOV-1996
DEFINITION zc64f09.r1 Soares fetal heart NBRHH19W Homo sapiens CDNA clone
327113 5' similar to WP:128D9.3 CE02068 ;
W30942
ACCESSION
NID G1311934
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini, Homiidae; Homo.
1 (bases 1 to 564)
REFERENCE
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfs, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT
Contact: Wilson RK
WashU-Merck EST Project

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 774 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 336.

FEATURES

Source

1..564
Location/Qualifiers
/organism="Homo sapiens"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCATCTGAAGTGGAGCGCCGATCTTTTATTTTATTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBHL19W."

/clone="327113"

/clone_lib="Soares fetal heart NBHL19W"

/sex="unknown"

/dev_stage="19 weeks"

/lab_host="DH10B (ampicillin resistant)"

<1..564

BASE COUNT 128 a 155 c 139 g 134 t 8 others

ORIGIN

Query Match 36.7%; Score 500; DB 1: Length 564;
Best Local Similarity 96.6%; Pred. No. 0.00e+00;
Matches 546; Conservative 0; Mismatches 12; Indels 7; Gaps 7;
Db 2 agcaccatcaagcctaccaccagggtttactgcaatgatgagatcaagtagccca 61
QY 462 AGCACCATCAAGCCTTACCACCGGAGGTTTACTGCAATGATGAGCATCAAGTACCCA 521
Db 62 ctgaactggtgagacaataatgacgtgtgtctgtgtccctggggatcgatgcc 121
QY 522 CTGAAAACCTGGTGGAGACAATAAATGACGCTGTCTGTGCGCGGGATCGTCATTGCC 581
Db 122 atctcgcatcatcacgggggaattctaccgcatctattacctgaagaatcgcggtcc 181
QY 582 ATCCTCGCATCATCAGGGGGGAATCTACCGGATCTATTACCTGAAGAAGTCGCGGTC- 640
Db 182 gacgattcagaacccctacgtggcagcactctataagcaagtgggtgcttctctttgg 241
QY 641 GACGATTCAAGACCCCTACGTGGCAGCACACTCTATAAGCAAGTGGGCTGCTTCCTTTGG 700
Db 242 ctgtgc-atcagcagctcttccagacattgccaagtgtccatagggcgctgctgcc 300
QY 701 CTGTGCATCAGCAGCTCTTTCACAGACATTCGCAAGTGTCCATAGGGGCGCTGCGTCC 760
Db 301 tcaactctgaagtgtcgaacccgtattctcagccagatcaactgctctgaagggtaca 360
QY 761 TCACCTCTTGA-GTGTCTGCAACCCGTATTTCAGCCAGATCAACTGCTCTGAAGGCTACA 819
Db 361 ttcaactacagatgcagaggtgatgacagcaaatccaggangccaggaaagtccttcn 420
QY 820 TTCAGAACTACAGATGCAGAGGTGATGACAGCAAGATCCAGGAAGCCAGGAATCTCTCT 879
Db 421 tctctggcaatgctctcttcacatgtatgactatgctggaattgtgtatnctnctgag 480
QY 880 TCCTGTGCCATGCTCTCTCTCCATGTACACTATGCTGTA-TTTGTGTCTATACCTGCGAG 938
Db 481 gcccgcttacttggcgagagcc-gnctactnccggccctctcgactnact-gatn 538
QY 939 GCCCGCTTCACTGGGAGAGAGCGCGCTGCTCGGGCCCCCTCTGCAAGTTCACCTTGATC 998

Db 539 atgatgggcctntacaacgggactg 563
QY 999 ATGATGGCCTTCTACA-CGGGACTG 1022

RESULT 2

ID HS942332 standard; RNA; EST; 564 BP.
AC W30942;
NI 91311934
DT 13-MAY-1996 (Rel. 47, Created)
DT 07-MAR-1997 (Rel. 51, Last updated, Version 2)
DE zc64f09.r1 Soares fetal heart NBHL19W Homo sapiens cDNA clone
DE 327113 5' similar to WP:T28D9.3 CE02068 ;
KW EST.
OS Homo sapiens (human)
OC Eukaryotes; mitochondria; eukaryotes; Metazoa; Chordata;
OC Vertebrata; Euthera; Primates; Catarrhini; Hominiidae; Homo.
RN 1..564
RP 1
RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,
RA Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,
RA Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F.,
RA Trevasis E., Waterston R., Williamson A., Wohldmann P., Wilson R.,
RT "The WashU-Merck EST Project";
RL Unpublished.
CC Contact: Wilton RK WashU-Merck EST Project Washington University
CC School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,
CC MO 63108 tel: 314 286 1800 Fax: 314 286 1810 Email:
CC est@watson.wustl.edu This clone is available royalty-free through
CC LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for
CC further information. Insert Length: 774 Std Error: 0.00 Seq primer:
CC mob.REGA+ET High quality sequence stop: 336.
FH Key Location/Qualifiers
FT source 1..564
FT /organism="Homo sapiens"
FT /note="Organ: heart; Vector: pT73D (Pharmacia) with a
FT modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
FT strand cDNA was primed with a Not I - oligo(dT) primer [5'
FT TGTTACCATCTGAAGTGGAGCGCCGATCTTTTATTTTATTTT 3']
FT double-stranded cDNA was size selected, ligated to Eco RI
FT adapters (Pharmacia), digested with Not I and cloned into
FT the Not I and Eco RI sites of a modified pT73 vector
FT (Pharmacia). Library went through one round of
FT normalization to a Cot = 5. Library constructed by M.Fatima
FT Bonaldo. This library was constructed from the same fetus
FT as the fetal lung library, Soares fetal heart NBHL19W."
FT /clone="327113"
FT /clone_lib="Soares fetal heart NBHL19W"
FT /sex="unknown"
FT /dev_stage="19 weeks"
FT /lab_host="DH10B (ampicillin resistant)"
FT mrna
SQ Sequence 564 BP; 128 A; 155 C; 139 G; 134 T; 8 other;

Query Match 36.7%; Score 500; DB 88; Length 564;
Best Local Similarity 96.6%; Pred. No. 0.00e+00;
Matches 546; Conservative 0; Mismatches 12; Indels 7; Gaps 7;

Db 2 agcaccatcaagcctaccaccagggtttactgcaatgatgagatcaagtagccca 61
QY 462 AGCACCATCAAGCCTTACCACCGGAGGTTTACTGCAATGATGAGCATCAAGTACCCA 521
Db 62 ctgaactggtgagacaataatgacgtgtgtctgtgtccctggggatcgatgcc 121
QY 522 CTGAAAACCTGGTGGAGACAATAAATGACGCTGTCTGTGCGCGGGATCGTCATTGCC 581
Db 122 atctcgcatcatcacgggggaattctaccgcatctattacctgaagaatcgcggtcc 181
QY 582 ATCCTCGCATCATCAGGGGGGAATCTACCGGATCTATTACCTGAAGAAGTCGCGGTC- 640

```

Db 182 gacgattcagaacccctacgtggcagcactctataagcaagtgggctcttctctcttgg 241
QY 641 GACGATTGAGAACCCCTACGTGGCAGCACTCTATAAGCAAGTGGCTGCTTCTCTTTGG 700

Db 242 ctgttc-atcagcagttttcagagacattgccaagtgtccatagggcgctcgctcc 300
QY 701 CBTGTCCTATCAGCCAGTCTTTCACAGACATTGCCAAAGTGTCCATAGGCGCCTCGCTCC 760

Db 301 tcaactttgaagtgtctgcgaacctgatttcagccagatcaactgctctgaaggtctaca 360
QY 761 TCACCTTCTTGA-GTGTCTGCAACCTGATTTCAGCCAGATCAACTGCTCTGAAGCTTACA 819

Db 361 ttcaagaactacagatgcagaggtgatagacgaaagtccaggangccagggaagtcctttn 420
QY 820 TTCAGAACTACAGATGCAGAGGTGATGACAGCAAAAGTCCAGGAAGCCAGGAAGTCTTCT 879

Db 421 tctctggcagctccctcttcacgtacacactatctctggaatttggctctatanctgag 480
QY 880 TCTCTGGCCATGCCCTTCTCCATGTACATGCTGTA-TTTGGTGTATACCTGTCAG 938

Db 481 gcccgttcaacttggcgaggagcc-gnctgctncgggcccctctcaggttnacct-gatn 538
QY 939 GCCCGCTTCACCTGGGAGGAGCCGCTGCTCCGCGCCCTCTCTGAGTTCACCTTGATC 998

Db 539 atgatgggctntacaacgggagctg 563
QY 999 ATGATGGGCTTCTACA-CGGGACTG 1022

RESULT 3 AA192756 441 bp mRNA EST 16-JAN-1997
LOCUS zq12c05.s1 Stratagene muscle 937209 Homo sapiens cDNA clone 629480
DEFINITION 3'
ACCESSION AA192756
NID 91782153
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 441)
AUTHORS Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,
Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,
Parsons J., Rifkin L., Rohlfing T., Tan F., Trevasakis E.,
Waterston R., Williamson A., Wohlmann P. and Wilson R.
TITLE WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: polyT not found
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 390.
Location/Qualifiers
1..441
/organism="Homo sapiens"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dt. Skeletal
muscle from patient with malignant hyperthermia. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGACGACGAG 3' -3' adaptor sequence: 5'
CTCGAGTCTTTTCTTTTCTTTT 3'"
/clone="629480"
/clone_lib="Stratagene muscle 937209"
/dev_stage="adult"
/lab_host="SOLR (kanamycin resistant)"

```

```

mRNA complement(<1..>441)
BASE COUNT 60 a 142 c 143 g 95 t 1 others
ORIGIN

Query Match 28.0%; Score 382; DB 15; Length 441;
Best Local Similarity 98.0%; Pred. No. 0.00e+00;
Matches 399; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

Db 1 ggcagatgagcagcaccgcgttgcgcgtccctccctcgtggttcttggtagcccgagc 60
Cp 408 GCGAGATGAGCAGCACCAGCCGCTTGTCCGCTCCCTCCCTCGGTTGTGTGAGCGCCGGC 349

Db 61 tgcgcgcgttcttctctccgggagcgtcttctgtctactgttagtttgcattggcgc 120
Cp 348 TCGCGCCGTTTCTGCTCTCCGGGAGCATCGCTTGTGCTACTTGTAGTTTGCATGGCGC 289

Db 121 tggctggcgccgagctc-cagggccccggaagcgtcccgcaacagcagcacacacca 179
Cp 288 TGGTTCGCGCGCAGACCTCTCCGCGCCCGCGAAGAGCTCCCGCAACAGCAGCACACCCA 229

Db 180 ggcgccgggtgcgc-tcctggccaggtcgtgcgagtagtggcggtcggcccggtc 238
Cp 228 GCGCGCCGGTTCGCTCTCGCGAGGCTGCTGCGAGTAGTGGCGGTCGCGCCGGGCTC 169

Db 239 cggcngcgcggttagtgcagccggggtgcttgcctccaaactgcagaaggtgtgtt 298
Cp 168 CGGCGCGCGGCTAGAGTGCAGCGCGGGGCTGCCTCCCACTGCAGAGAGGTGTGT 109

Db 299 tctgctcctcctgcgcctctctctctctctctctcctcctcctcctcctcctcctc 358
Cp 108 TTCTGCTCTCTCTGCGCGCTCTGCTTTCTCTGCTCAACCTCAATCTCTTAAAGTCCA 49

Db 359 aggggaagcagcagatcccgagcagaaacttttgcagagctcgc 405
Cp 48 AGGGGAAGAGAGCAGATCCCGCAGCAAAACTTTTTCAGAGCTGCGC 2

RESULT 4
ID HS733324 standard; RNA; EST; 403 BP.
AC W02733;
NI 91274782
DT 25-APR-1996 (Rel. 47, Created)
DE zc64f09.s1 Soares fetal heart NbHH19W Homo sapiens cDNA clone
DE 327113 3'.
KW EST.
OS Homo sapiens (human)
OC Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
OC Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP 1-403
RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,
RA Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,
RA Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F.,
RA Trevasakis E., Waterston R., Williamson A., Wohlmann P., Wilson R.;
RA "The WashU-Merck EST Project";
RL Unpublished.
CC Contact: Wilson RK WashU-Merck EST Project Washington University
CC School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,
CC MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:
CC est@watson.wustl.edu This clone is available royalty-free through
CC LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for
CC further information. Insert Length: 774 Std Error: 0.00 Seq primer:
CC mob.RBGA+ET High quality sequence stop: 312.
Key Location/Qualifiers
source
1..403
/organism="Homo sapiens"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGCATCTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI

```


REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	(bases 1 to 446)			
1	Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.	WashU-Merck EST Project		Unpublished (1995)
1	Contact: Wilson RK			
1	WashU-Merck EST Project			
1	Washington University School of Medicine			
1	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108			
1	Tel: 314 286 1800			
1	Fax: 314 286 1810			
1	Email: est@watson.wustl.edu			
1	This clone is available royalty-free through LLNL ; contact the			
1	IMAGE Consortium (info@image.llnl.gov) for further information.			
1	Seq primer: -40M13 fwd. from AmerSham			
1	High quality sequence stop: 286.			
1	Location/Qualifiers			
1	1...446			
1	/organism="Homo sapiens"			
1	/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:			
1	XhoI; Cloned unidirectionally. Primer: Oligo dT. Hela S3			
1	epithelioid carcinoma cells grown to semi-confluency			
1	without induction. Average insert size: 1.5 kb; Uni-ZAP XR			
1	Vector. -5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3'			
1	adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'			
1	/clone="613099"			
1	/clone_lib="Stratagene HeLa cell s3 937216"			
1	/sex="female"			
1	/dev_stage="HeLa S3 cell line"			
1	/lab_host="SOLR (kanamycin resistant)"			
1	complement(<1..>446)			
1	111 a 103 c 125 g 101 t 6 others			
1	Query Match 24.7%; Score 336; DB 12; Length 446;			
1	Best Local Similarity 99.4%; Pred. No. 0.00e+00;			
1	Matches 342; Conservative 0; Mismatches 1; Indels 1; Gaps 1;			
1	Db 1 gtttttcctacattctactgtctgtgatgagattggagagcagcaagaactgtgtgcagc 60			
1	Cp 1327 GTTTTTCCTTACATTCCTACTGTCTGTATGAGATTGGAGCAGCAGCAAGAACTTGTGTGTCAGC 1268			
1	Db 61 agtcattttcaaaanacagctcaggaggtgggtggcgacctacatcatgttgggtgattg 120			
1	Cp 1267 AGTCATTTTACAAAACAGCTCAGAGGTGGTGGCACCCTACATCATGTGTGGTGATTG 1208			
1	Db 121 ttctgtcaataatgtccacaggtgaagagatttccctccgagatcagcgggcaggcagg 180			
1	Cp 1207 TTCTGTCAATATATGTCCACAGGTGAAGAGATTTCCTCCGATGAGCGGGCAGGCAGG 1148			
1	Db 181 gagagcgtctcttagtcttgaagaggtccagacacgaagaaactatgcagcagccacc 240			
1	Cp 1147 GAGAGCGTCGCTTAGTCTTGAAGAGTCAGACACGAGAAAACATATGCAGCAGGCCACC 1088			
1	Db 241 agggctccttgagcaaatcctgccagaacatcactggatgggtgttctgtgtgtatcagc 300			
1	Cp 1087 AGGCTCCTTGTAGCAAACTCCTGCCAGAACATCACTGGGATGGTGTGTGTCTGTATACG 1028			
1	Db 301 cgagacagtcctcgttagaagggccatcatgatcaagtgaact 344			
1	Cp 1027 CGAGACAGTCCC-GTGTAGAGGCCCATCATGATCAAGGTGAACT 985			
1	RESULT 7			
1	LOCUS W39811 402 bp mRNA EST 05-FEB-1997			
1	DEFINITION 307 Mouse VM cDNA library Mus musculus cDNA clone smp22 1.22.			
1	ACCESSION W39811			
1	NID g1816977			

KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	1 (bases 1 to 402)
AUTHORS	Stewart,G.J., Savioz,A. and Davies,R.W.
TITLE	Sequence analysis of 497 mouse brain ESTs expressed in the substantia nigra
JOURNAL	Genomics 39, 147-153 (1997)
COMMENT	Contact: Davies,R.W. Robertson Laboratory of Biotechnology Institute of Biomedical and Life Sciences, Division of Molecular Genetics Robertson Building, 54 Dumbarton Road, Glasgow G11 6NU Tel: 44 141 330 5102 Fax: 44 141 330 5102/4878 Email: gbga21eudcf.gla.ac.uk Seq primer: T7/T3alpha High quality sequence stop: 422. Location/Qualifiers 1..402 /organism="Mus musculus" /note="Organ: brain; Vector: pSPORT1; Mouse Ventral Midbrain directional cDNA library in pSPORT1. The library was created by subtractive hybridisation using VM mRNA as the target nucleic acid population and total cerebellar ss cDNA as the driver nucleic acid population. The clones are 3' directed." /clone="smp22 1.22" /clone_lib="Mouse VM cDNA library" /tissue_type="ventral midbrain" <1..>402
BASE COUNT	69 a 131 c 101 g 101 t
ORIGIN	
Query Match	18.4%; Score 250; DB 27; Length 402;
Best Local Similarity	84.4%; Pred. No. 0.00e+00;
Matches	313; Conservative 0; Mismatches 57; Indels 1; Gaps 1;
Db	2 ggccagcgcctctctccatgttccactatgctgtatctgtgtgtctacaccttcaggccgcg 61
Qy	885 GGCCATGCCCTCTTCATCGTACACATGCTGTATTGGTGCTATACCTCGAGCCGCG 944
Db	62 ttcacctggcggggggccgactgcgcgccctctgcagttgcactttgctcatgatg 121
Qy	945 TTCACTTTGGGAGGAGCGCGCTCTCGGCGCCCTCCTCGAGTTTACCTTGATCATGTG 1004
Db	122 gccttctacagcgatgtcacgggtatctgactacaagcatcatcctagcgatgctcg 181
Qy	1005 GCCTTCTACAGGAGACTGTCTCGCGTATACAGCACCAAGACCATTCCAGTGTGTTCTG 1064
Db	182 gcaggattggcccaagagcctctgtgtggcctgctgcatagtgtcttcgttccgacctc 241
Qy	1065 GCAGATTGCTCAAGGAGCCCTGGTGCCCTGCTGCATAGTTTCTTCGTCCTGACCTC 1124
Db	242 ttcagactaagaagagcctctcactgcccgcctctgcgatcaggaggagatcctgtct 301
Qy	1125 TTCAAGACTAAGACGAGCTCTCCCTGCTGCCCTGCTATCCGGAAGAAATCCTTTCA 1184
Db	302 cccgtgacatcatcgccagggaacaatcaccataccatgtgttagatgtcggcgctccc 361
Qy	1185 CCTGTGACATATTATGACAGG-AACAATCACCAACAATGATGTAGTGTGCCACCCACCTC 1243
Db	362 cggagcgcttt 372
Qy	1244 CTGAGCTGTTT 1254
RESULT	8 W39816 337 bp mRNA EST 05-FEB-1997
LOCUS	

RESULT	8	w39816	337 bp	EST	05-FEB-1997
LOCUS				mrna	

DEFINITION	309 Mouse VM cDNA library Mus musculus cDNA clone smp22 1.25.
ACCESSION	W39816
NID	91816979
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	1 (bases 1 to 337)
AUTHORS	Stewart,G.J., Savioz,A. and Davies,R.W.
TITLE	Sequence analysis of 497 mouse brain ESTs expressed in the
JOURNAL	substantia nigra
COMMENT	Genomics 39, 147-153 (1997)
CONTACT	Contact: Davies,R.W.
INSTITUTE	Robertson Laboratory of Biotechnology
GENETICS	Institute of Biomedical and Life Sciences, Division of Molecular
TELEPHONE	Robertson Building, 54 Dumbarton Road, Glasgow G11 6NU
FAX	Tel: 44 141 330 5102
EMAIL	Fax: 44 141 330 5102/4878
SEQ PRIMER	Email: gbga21@udcf.gla.ac.uk
SEQ PRIMER	Seq primer: T7/T3alpha
High quality sequence stop: 337.	
FEATURES	Location/Qualifiers
source	1..337
	/organism="Mus musculus"
	/note="Organ: brain; Vector: pSPORT1; Mouse Ventral
	Midbrain directional cDNA library in pSPORT1. The library
	was created by subtractive hybridisation using VM mRNA as
	the target nucleic acid population and total cerebellar ss
	cDNA as the driver nucleic acid population. The clones are
	3' directed."
	/clone="smp22 1.25"
	/clone_lib="Mouse VM cDNA library"
	/tissue_type="ventral midbrain"
mRNA	<1..>337
BASE COUNT	59 a 112 c 83 g 83 t
ORIGIN	
Query Match	17.9%; Score 244; DB 27; Length 337;
Best Local Similarity	87.4%; Pred. No. 0.00e+00;
Matches	291; Conservative 0; Mismatches 41; Indels 1; Gaps 1;
Db	3 gccacgctctctccatgtttcaactatctgtatctgtgtctaccctcagggccgc 62
Qy	
	885 GGCCATGCCCTCTCTCCATGTACACTATGCTGTATTGGTGTATATACCTGCAGGCCGC 944
Db	63 ttcaacctggcgggggccgcgactgtctcgcctctcctcagttcactttgctcatgtg 122
Qy	
	945 TTCACTTTGGCGAGGAGCCGCCCTGCTCGGCCCTCTCTGCAGTTCACCTTGATCATGATG 1004
Db	123 gcctttacacgggattgttcacgggtatctgactacagcatcatcctagcgtgtcctg 182
Qy	
	1005 GCCTTCTACACGGGACTGTCTCGGTATCAGACCACACAGCACCATCCAGTGATGTTCTG 1064
Db	183 gcaggatttcccaagagctctgttggtcctgtcgtcatagtgtcttcgttcgcagctc 242
Qy	
	1065 GCAGGATTGCTCAAGGAGCCCTGGTGGCCTGCTGTCATAGTTTCTTCGTGCTGCACCTC 1124
Db	243 ttcaagactaagaagacccttcaactcgtccgcctcgtcgtcagggagagatcctctct 302
Qy	
	1125 TTCAAGACTTAGACGACGCTCTCCTCGCTGCCCTCTATCCGGAAGGAATCCTTTCA 1184
Db	303 cccgtgacatcatcgtccagggaacaatcacca 335
Qy	
	1185 CCTGTGACATTATTGACAGG-AAACAATCACCA 1216
RESULT	9 G13351 546 bp DNA STS 20-DEC-1995
LOCUS	human STS WI-12145.
DEFINITION	

ACCESSION	GL13351
NID	gl127460
KEYWORDS	STS sequence; primer; sequence tagged site.
SOURCE	human STSs derived from sequences in dbEST and the Unigene collection.
ORGANISM	Homo sapiens
REFERENCE	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 546)
TITLE	Hudson, T.
JOURNAL	Whitehead Institute/MIT Center for Genome Research; Physically Mapped STSs
COMMENT	Unpublished (1995)
	Contact: Thomas Hudson
	Whitehead Institute/MIT Center for Genome Research
	Whitehead Institute for Biomedical Research
	9 Cambridge Center, Cambridge MA 02142 USA
	Tel: 617 252 1900
	Fax: 617 252 1902
	Email: thudson@genome.wi.mit.edu
	Primer A: AAAATCAGTCGGGCAAAAGT
	Primer B: TATTGACAGGAACAATCACCAC
	STS size: 150
	PCR Profile:
	Presoak:
	Annealing: 56 degrees C
	Polymerization:
	PCR Cycles: 35
	Thermal Cycler:
	Protocol:
	Template: 10 ng
	Primer: each 5 pM
	dNTPs: each 4 nM
	Taq Polymerase: 0.025 units/ul
	Total Vol: 20 ul
	Buffer:
	MgCl2: 1.5 mM
	KCl: 50 mM
	Tris-HCl: 10 mM
	pH: 9.3
FEATURES	Derived from dbEST (genbank accession T71976).
source	Location/Qualifiers
	1..546
STS	/organism="Homo sapiens"
	1..150
primer_bind	/map="981.7 cR from top of Chr2 linkage group"
	1..20
primer_bind	/map="981.7 cR from top of Chr2 linkage group"
	complement(129..150)
	/map="981.7 cR from top of Chr2 linkage group"
BASE COUNT	148 a 109 c 147 g 128 t 14 others
ORIGIN	
	Query Match 17.8%; Score 243; DB 40; Length 546;
	Best Local Similarity 98.4%; Pred. No. 0.00e+00;
	Matches 247; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Db	1 aaatcagtcgggcaaaagttttccctacattctctgtatgagatggagagcag 60
Cp	1345 AAAATCAGTCGGGCAAAAGT CCCTACATTCTACTGTGATGTGAGAGCAG 1286
Db	61 caagaacttctgtccacgagtcattttacaaaaaacagctcaggaggtgggtggcacctac 120
Cp	1285 CAAGAACCTTGTGTGACAGCGTCATTTTACAAAAACAGCTCAGGAGGTGGGTGGGACCTTAC 1226
Db	121 atcatgttgtgtgattgttctctgtcaataatgtccacaggtgaaaggatttctctccgg 180


```

Cp 1225 ATCAATGTTGGTGAATTCCTCTGTCATAATGTCACAGGTGAAGGATTTCCTTCGG 1166
|||||
Db 181 atagcagggcaggcaggagcgcctcttagtcttgaagaggtcagacacgaagaaa 240
|||||
Cp 1165 ATAGCAGGGCAGGCGAGGAGCGTCTCTAGTCTTGAAGAGGTGACAGACGAGAA 1106
|||||
Db 241 acctagaagca 251
|||
Cp 1105 ACTATGCAGCA 1095
|||||

RESULT 10
LOCUS AAL126964 421 bp mRNA EST 26-NOV-1996
DEFINITION z187e03.s1 Stratagene colon (#937204) Homo sapiens CDNA clone
511612 3'.
ACCESSION AAL126964
NID G1686390
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euthera; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 421)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevasakis, E.,
Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
WashU-Merck EST Project
WashU-Merck EST Project
Unpublished (1995)
TITLE Contact: Wilson RK
JOURNAL WashU-Merck EST Project
COMMENT Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 391.
FEATURES
source
Location/Qualifiers
1..421
/organism="Homo sapiens"
/note="Organ: colon; Vector: pBluescript SK-; Site.1:
EcoRI; Site.2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. T-84 clonic epithelial cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGGCAGG 3' -3' adaptor sequence: 5'
CTCGAGTCTTTTCTTTTCTTTT 3'"
/clone="511612"
/clone.lib="Stratagene colon (#937204)"
/lab_host="SOER cells (kanamycin resistant)"
complement(1..>421)
mRNA 133 a 78 c 120 g 90 t
BASE COUNT
ORIGIN
Query Match 16.5%; Score 225; DB 8; Length 421;
Best Local Similarity 98.3%; Pred. No. 0.00e+00;
Matches 229; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 1 gttttccctacattctactgtctgatgattggagagcagcaaacctgctcagc 60
|||||
Cp 1327 GTTTTCCCTACATCTCTACTGTCGTATGATGGATGGAGAGCAGCAGACTTGCTGTGACG 1268
|||||
Db 61 agtcattttacaaaaaacagctcaggaggtgggtggcaccctacatcatgttgggtgattg 120
|||||
Cp 1267 AGTCATTTTACAAAAACAGCTCAGGAGGTGGGTGGCACCACATCATCATGTGTGGTGATTG 1208
|||||
Db 121 ttccctgcaataatgtccacaggtgaaaggatttccctccggatgacggggcagggcagg 180
|||||
Cp 1207 TTCCCTGTCATAATGTCCACAGGTGAAAGGATTTCCTTCCGGATAGCAGGGGCGAGG 1148
|||||

```

```

Db 181 gagagcgtcgtcttagtcttgaagaggtcagacacgaagaaaacctagaagca 233
|||||
Cp 1147 GAGAGCGTGGCTTCTAGTCTTGAAGAGGTGACAGACAGGAAGAACTATGTCAGCA 1095
|||||

RESULT 11
LOCUS W23530 374 bp mRNA EST 25-NOV-1996
DEFINITION zc71h10.s1 Soares fetal heart NBHH19W Homo sapiens CDNA clone
327811 3'.
ACCESSION W23530
NID G1300561
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euthera; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 374)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
TITLE Contact: Wilson RK
JOURNAL WashU-Merck EST Project
COMMENT Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 716 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 262.
FEATURES
source
Location/Qualifiers
1..374
/organism="Homo sapiens"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand CDNA was primed with a Not I - Oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot - 5. Library was constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBHH19W."
/clone="327811"
/clone.lib="Soares fetal heart NBHH19W"
/sex="unknown"
/dev_host="DH10B (ampicillin resistant)"
lab_host="DH10B (ampicillin resistant)"
complement(1..>374)
mRNA 102 a 91 c 73 g 101 t 7 others
BASE COUNT
ORIGIN
Query Match 15.5%; Score 211; DB 1; Length 374;
Best Local Similarity 96.5%; Pred. No. 4.06e-296;
Matches 223; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

```

```

Db 38 ctacagcagccgtctccctgctgctccctgtctatccggaag-aaatcctttcactgtgg 96
|||||
Qy 1132 CTAGACGACGCTCTCCCTGCTGCCCTGCTATCCGGAAGGAATCTTTCCACTGTGG 1191
|||||
Db 97 acattattgacaggaacaatcaccacacacatgatgtagggtgccacccacctcctgagtg 156
|||||
Qy 1192 ACATTATTGACAGGAACAATCACCACACATGATGATGTAGGTGCCACCCACCTCCTGAGCTG 1251
|||||

```

```

Db 157 tttttaaagtactgctgacagcaagttcttgctgctctccaatctcatcagcagta 216
      |||||||
QY 1252 TTTTGTAAATGACTGCTGACAGCAAGTCTTGTGCTCTCCAATCTCATCAGACAGTA 1311
      |||||||

Db 217 gaatgtagggaataacttttgcgcgactgatttttaaaaggnnaaaaaa 267
      |||||||
QY 1312 GAATGTAGGGAATAACTTTTGCCTGACTGATTTTAAAAAATAAAAAA 1362

RESULT 12
LOCUS HS530338 standard; RNA; EST; 374 BP.
AC W23530;
NI g1300561
DT 09-MAY-1996 (Rel. 47, Created)
DE zc71h10.s1 Soares fetal heart NBHH19W Homo sapiens cDNA clone
DE 327811 3'.
KW EST.
OS Homo sapiens (human)
OC Eukaryotae; mitochondria eukaryotes; Metazoa; Chordata;
OC Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP 1-374
RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,
RA Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,
RA Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F.,
RA Trevasakis E., Waterston R., Williamson A., Wohlmann P., Wilson R.;
RT "The WashU-Merck EST Project";
RL Unpublished.
CC Contact: Wilson RK WashU-Merck EST Project Washington University
CC School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,
CC MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:
CC estevaton.wustl.edu This clone is available royalty-free through
CC LNL ; contact the IMAGE Consortium (infoimage.llnl.gov) for
CC further information. Insert Length: 716 Std Error: 0.00 Seq primer:
CC mob.REGA+ET High quality sequence stop: 262.
FH Key
FH Location/Qualifiers
FT 1..374
FT /organism="Homo sapiens"
FT /note="Organ: heart; Vector: pTT73D (Pharmacia) with a
FT modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
FT strand cDNA was primed with a Not I - oligo(dT) primer [5',
FT TGTACCAATCTGAAGTGGAGCGCGCATCTTTTATTTT 3']
FT double-stranded cDNAs size selected, ligated to Eco RI
FT adapters (Pharmacia), digested with Not I and cloned into
FT the Not I and Eco RI sites of a modified pT73 vector
FT (Pharmacia). Library went through one round of
FT normalization to a Cot = 5. Library constructed by M.Fatim
FT a
FT Bernaldo. This library was constructed from the same fetus
FT as the fetal lung library, Soares fetal lung NBHH19W."
FT /clone="327811"
FT /clone_lib="Soares fetal heart NBHH19W"
FT /sex="unknown"
FT /dev_stage="19 weeks"
FT /lab_host="DH10B (ampicillin resistant)"
FT complement(<1..>374)
FT mRNA
FT Sequence 374 BP; 102 A; 91 C; 73 G; 101 T; 7 other;

Query Match 15.5%; Score 211; DB 87; Length 374;
Best Local Similarity 96.5%; Pred. No. 4.06e-296;
Matches 223; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

Db 38 ctacagcgcgtctccctgcgcctctctatccggaag-aatcttccactgtg 96
      |||
QY 1132 CTAAGAGGAGCGCTCTCCCTGCCCTGCTATCCGGAAGAAATCTTCACTGTGG 1191

Db 97 acattattgacaggaacaaatcacacaacatgatgtagtgccaccacctctgagct 156
      |||||||
QY 1192 ACATATTAGGAGGAACAATCACACAAATGATAGGTGGCCACCACCTCTGAGCTG 1251

```

```

Db 157 tttttaaagtactgctgacagcaagttcttgctgctctccaatctcatcagcagta 216
      |||||||
QY 1252 TTTTGTAAATGACTGCTGACAGCAAGTCTTGTGCTCTCCAATCTCATCAGACAGTA 1311
      |||||||

Db 217 gaatgtagggaataacttttgcgcgactgatttttaaaaggnnaaaaaa 267
      |||||||
QY 1312 GAATGTAGGGAATAACTTTTGCCTGACTGATTTTAAAAAATAAAAAA 1362

RESULT 13
LOCUS W39815 242 bp mRNA EST 05-FEB-1997
DEFINITION 308 Mouse VM cDNA library Mus musculus cDNA clone spmp22 1.23.
ACCESSION W39815
NID g1816978
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotae; mitochondria eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Mus.
REFERENCE 1 (bases 1 to 242)
AUTHORS Stewart,G.J., Savioz,A. and Davies,R.W.
TITLE Sequence analysis of 497 mouse brain ESTs expressed in the
JOURNAL Genomics 39, 147-153 (1997)
COMMENT Contact: Davies,R.W.
Robertson Laboratory of Biotechnology
Institute of Biomedical and Life Sciences, Division of Molecular
Genetics
Robertson Building, 54 Dumbarton Road, Glasgow G11 6NU
Tel: 44 141 330 5102
Fax: 44 141 330 5102/4878
Email: gba21@udcf.gla.ac.uk
Seq primer: T7/T3alpha
High quality sequence stop: 2.
FEATURES
source
Location/Qualifiers
1..242
/organism="Mus musculus"
/note="Organ: brain; Vector: pSPORT1; Mouse Ventral
Midbrain directional cDNA library in pSPORT1. The library
was created by subtractive hybridisation using VM mRNA as
the target nucleic acid population and total cerebellar ss
cDNA as the driver nucleic acid population. The clones are
3' directed."
/clone="spmp22 1.23"
/clone_lib="Mouse VM cDNA library"
/tissue_type="ventral midbrain"
<1..>242
mRNA
BASE COUNT 34 a 82 c 60 g 66 t
ORIGIN
Query Match 14.0%; Score 190; DB 27; Length 242;
Best Local Similarity 89.6%; Pred. No. 2.04e-260;
Matches 215; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Db 3 gaccacgcctctctccatgttctacatgctgctatctgctctctacacctcaggccgcg 52
      |||||
QY 885 GGCCATGCCCTCTCTCCATGACACTATGCTGTATTTGGTGTATATACCTGCAGGCCCGC 944

Db 63 ttacacctggcggggccgcgactgctccgccccctcctcagttcacttgcctcatgatg 122
      |||||
QY 945 TTCACCTGGCGAGGAGCGCCCTGCTCCGGCCCCCTCTCGACGTTTACCTTGATCATGATG 1004

Db 123 gcocttacacgggagtggttcacgggtatctgactacagcatcatcctcagctgctctg 182
      |||||
QY 1005 GCCTTCTACACGGGAGTGTCTCGCGTATCAGACCACCAAGCACCACCTCCAGTGTCTCG 1064

Db 183 gcaggatttgcccaaggagactgctggcgctgctgcacatagttcttctggtccgacctc 242
      |||||
QY 1065 GCAGGATTTGCTCAAGGAGCCCTGTGGCTGCTGCATAGTTTCTTCTGCTGACCTC 1124

```

```
RESULT 14
ID HS1146241 standard; RNA; EST; 236 BP.
AC AA182434;
NI g1766132
DT 24-FEB-1997 (Rel. 51, Created)
DE 2p52f10.r1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone
DE 613099 5' similar to TR:G1161100 G1161100 HYDROGEN
DE PEROXIDE-INDUCIBLE PROTEIN ;
KW EST.
OS Homo sapiens (human)
OC Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
OC Vertebrata; Euthera; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP 1-236
RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,
RA Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,
RA Parsons J., Rifkin L., Rohlfing T., Tan F., Trevaskis E.,
RA Waterston R., Williamson A., Wohlmann P., Wilson R.;
RA "WashU-Merck EST Project";
RL Unpublished.
RC Contact: Wilson RK WashU-Merck EST Project Washington University
CC School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,
CC MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:
CC estevaton.wustl.edu This clone is available royalty-free through
CC LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for
CC further information. Seq primer: -28M13 rev2 from Amersham High
CC quality sequence stop: 91
CC Key Location/Qualifiers
FH source
FH 1..236
FH /organism="Homo sapiens"
FH /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
FH XhoI; Cloned unidirectionally. Primer: Oligo dt. HeLa S3
FH epithelioid carcinoma cells grown to semi-confluency
FH without induction. Average insert size: 1.5 kb; Uni-ZAP XR
FH Vector. -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3'
FH adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
FH /clone="613099"
FH /clone.lib="Stratagene HeLa cell s3 937216"
FH /sex="female"
FH /dev_stage="HeLa S3 cell line"
FH /lab_host="SOLR (kanamycin resistant)"
FH <1..>236
FH mrna
FH Sequence 236 BP; 60 A; 59 C; 52 G; 56 T; 9 other;
FH
FH Query Match 13.1%; Score 179; DB 81; Length 236;
FH Best Local Similarity 91.1%; Pred. No. 8.22e-242;
FH Matches 216; Conservative 0; Mismatches 16; Indels 5; Gaps 5;
Db 1 ttcaagaacccctacgtggcagcactctataagcaagtggtgctgctnctcttggctgt- 59
Qy 646 TTCAGAACCCCTACGTGGCAGCAGCTCTATAGCAAGTGGGCTGCTCTCTTTGGCTGTG 705
Db 60 catcaagccagcttctcacagacattgccaaagtgtccataggcgctgctnctcaact 119
Qy 706 CCATCAGCCAGCTCTTTCACAGACATTGCCAAAGTGTCCATAGGGCGCTGCTCCTCACT 765
Db 120 tcntgagtgctnaaccctgatttnaagccagagtaacttgctctgaaggctacatnca 179
Qy 766 TCTTGAGTGTCTGCAACCTGATTTCACATCAACT-GCTCTGAAGGCTACATTCAC 823
Db 180 gaactacagatgnagaggtgatgcagcaagtcacaggaagcgaaggaagtccttc 236
Qy 824 GAACT-ACAGATGCAGAGGTGTATGACAGCAAAAGTCCAGGAAGCCA-GGAAGTCCTTC 878
Search completed: Thu Nov 6 12:23:06 1997
Job time : 443 secs.
```

```
g1766132
EST.
human.
Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euthera; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 236)
Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,
Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,
Parsons J., Rifkin L., Rohlfing T., Tan F., Trevaskis E.,
Waterston R., Williamson A., Wohlmann P., and Wilson R.
WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estevaton.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 91.
Location/Qualifiers
1..236
/organism="Homo sapiens"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dt. HeLa S3
epithelioid carcinoma cells grown to semi-confluency
without induction. Average insert size: 1.5 kb; Uni-ZAP XR
Vector. -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3'
adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
/clone="613099"
/clone.lib="Stratagene HeLa cell s3 937216"
/sex="female"
/dev_stage="HeLa S3 cell line"
/lab_host="SOLR (kanamycin resistant)"
<1..>236
60 a 59 c 52 g 56 t 9 others
mrna
BASE COUNT
ORIGIN
Query Match 13.1%; Score 179; DB 12; Length 236;
Best Local Similarity 91.1%; Pred. No. 8.22e-242;
Matches 216; Conservative 0; Mismatches 16; Indels 5; Gaps 5;
Db 1 ttcaagaacccctacgtggcagcactctataagcaagtggtgctgctnctcttggctgt- 59
Qy 646 TTCAGAACCCCTACGTGGCAGCAGCTCTATAGCAAGTGGGCTGCTCTCTTTGGCTGTG 705
Db 60 catcaagccagcttctcacagacattgccaaagtgtccataggcgctgctnctcaact 119
Qy 706 CCATCAGCCAGCTCTTTCACAGACATTGCCAAAGTGTCCATAGGGCGCTGCTCCTCACT 765
Db 120 tcntgagtgctnaaccctgatttnaagccagagtaacttgctctgaaggctacatnca 179
Qy 766 TCTTGAGTGTCTGCAACCTGATTTCACATCAACT-GCTCTGAAGGCTACATTCAC 823
Db 180 gaactacagatgnagaggtgatgcagcaagtcacaggaagcgaaggaagtccttc 236
Qy 824 GAACT-ACAGATGCAGAGGTGTATGACAGCAAAAGTCCAGGAAGCCA-GGAAGTCCTTC 878
Search completed: Thu Nov 6 12:23:06 1997
Job time : 443 secs.
```

This Page Blank (uspto)

W P S R C H

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by Intelligenetics, Inc.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Nov 4 10:27:48 1997; MasPar time 7.48 seconds
Tabular output not generated.
511.387 Million cell updates/sec

Title: >US-08-842-827-6
Description: (1-311) from US08842827.pap
Perfect Score: 2326
Sequence: 1 MQNYKYDKAIVPESKGGSP.....RKEILSPVDIIDRNHHNMM 311

Scoring table: PAM 150
Gap 11

Searched: 101610 seqs, 12294212 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq28
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21

Statistics: Mean 33.923; Variance 140.094; scale 0.242

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	111	4.8	352	20	W01663	p(H218), G-protein co
2	111	4.8	352	11	R58712	Fragment of receptor
3	110	4.7	251	21	W20277	H. pylori surface or
4	96	4.1	111	21	R57067	Human serotonergic
5	95	4.1	492	2	R11360	Glucose Transporter p
6	95	4.1	836	17	R87153	Alternatively spliced
7	95	4.1	836	11	R58912	Product of alternativ
8	95	4.1	904	17	R87147	Protocadherin clone 4
9	95	4.1	904	11	R58907	Human protocadherin-4
10	93	4.0	370	11	R57066	Murine serotonergic
11	90	3.9	583	20	W06891	Human activated leuko
12	90	3.9	583	18	R97230	Stem cell marker HCAP
13	89	3.8	263	4	R23424	Human MIP.
14	89	3.8	370	12	R58686	Rat MR22 serotonin re
15	88	3.8	588	19	W05297	Esterase secretory pr
16	86	3.7	88	16	R90991	Mouse Mat-8 polypepti
17	87	3.7	121	21	W20696	H. pylori secreted or
18	87	3.7	216	15	R75902	Human olfactory recep
19	87	3.7	357	9	R45847	Murine 5HT5a serotoni
20	86	3.7	357	12	R58685	Rat REC17 serotonin r

21	85	3.7	357	9	R45848	Human 5HT5a serotonin	6.70e+01
22	87	3.7	377	21	W20661	H. pylori transmembra	4.83e+01
23	85	3.7	829	16	R65578	Autotaxin derived fro	6.70e+01
24	85	3.7	849	16	R86595	N-tera 2D1 autotaxin	6.70e+01
25	85	3.7	915	16	R86596	A2058 autotaxin prote	6.70e+01
26	84	3.6	229	13	R74037	MHC polypeptide HLA D	7.87e+01
27	83	3.6	324	20	W11860	Human proteosome subu	9.25e+01
28	83	3.6	369	7	R39261	Human somatostatin re	9.25e+01
29	83	3.6	369	7	R39262	Murine somatostatin r	9.25e+01
30	83	3.6	369	18	R97269	Human somatostatin r	9.25e+01
31	83	3.6	455	3	R12362	Octopus rhodopsin mem	9.25e+01
32	83	3.6	570	18	R97231	Stem cell marker HCAP	9.25e+01
33	84	3.6	724	5	R27648	Human calcium channel	9.25e+01
34	84	3.6	1138	2	R06461	BPGS1245 protoxin.	7.87e+01
35	84	3.6	1138	9	R46226	Bacillus thuringiens	7.87e+01
36	84	3.6	1138	7	R37214	B.t. toxin HD867	7.87e+01
37	84	3.6	1138	9	R46225	Bacillus thuringiens	7.87e+01
38	84	3.6	1138	7	R37213	B.t. toxin HD511.	7.87e+01
39	83	3.6	2251	14	R71009	Human neuronal calciu	9.25e+01
40	83	3.6	2270	14	R71010	Human neuronal calciu	9.25e+01
41	83	3.6	2270	13	R69604	Calcium channel alpha	9.25e+01
42	83	3.6	2458	2	R07640	Deduced protein seque	9.25e+01
43	83	3.6	2458	2	R04031	Full length T4 encode	9.25e+01
44	83	3.6	4472	19	R97245	Virulence gene cluste	9.25e+01
45	82	3.5	368	1	R06606	Lipase encoded by rec	1.09e+02

ALIGNMENTS

RESULT 1
ID W01663 standard; Protein; 352 AA.
AC W01663;
DT 01-APR-1997 (first entry)
DE p(H218), G-protein coupled receptor.
KW p(H218); G-protein coupled receptor; cell differentiation; proliferation;
KW proline directed kinase; cell division; growth factor response; rat-edg;
KW therapy; diagnosis.
OS Rattus rattus.
FT Key Location/Qualifiers
FT Modified_site 19
FT /note= "predicted N-glycosylation site"
FT Binding_site 222..225
FT /note= "possible nucleotide binding site"
FT Binding_site 309..315
FT /note= "possible nucleotide binding site".
FT Misc_difference 342..343
FT /note= "consensus sequence for proline directed
FT kinases"
FT US5585476-A.
PD 17-DEC-1996.
PF 15-FEB-1994; 196989
PR 15-FEB-1994; US-196989.
PA (MACL/) MACLENNAN A J.
PI MacIennan AJ;
DR WPI; 97-051235/05.
DR N-PSDB; T58505.
PT DNA encoding rat protein p(H218) - associated with cell
PT proliferation and/or differentiation
PS Claim 1; Column 17-20; 3pp; English.
CC This sequence is a novel rat protein p(H218), a member of the G-protein
CC coupled receptor (GPR) superfamily of proteins. Several features of
CC p(H218) are common to all other GPRs, including: (i) seven regions of
CC hydrophobicity which are predicted to act as membrane spanning domains;
CC (ii) a consensus sequence for N-linked glycosylation in its predicted
CC N-terminal extracellular domain, and (iii) a conserved cysteine residue
CC and several serine and threonine residues in its predicted C-terminal
CC domain. In addition, p(H218) contains many other residues which are
CC highly conserved among most GPRs. However, p(H218) is distinct from these
CC GPRs in that it does not contain certain highly conserved residues.
CC Perhaps most notable are the aspartate and tyrosine residues at the
CC cytoplasmic end of the third transmembrane domain, and the cysteine
CC residue at the extracellular end of the same transmembrane domain. The
CC amino acid similarity between p(H218) and p(rat-edg) (W01664) suggests

PT increase the productivity and extracellular secretion of esterases
 PS Claim 3; Page 15-17; 29pp; English.
 CC 3 Different proteins (W05297-99) participate in the mechanism of
 CC secretion of esterase by Serratia marcescens Sr41. They are
 CC encoded by an esterase secretory gene (T39862) that includes 3
 CC open reading frames. The proteins can be produced by transformed
 CC host cells, pref. S. marcescens or E. coli, carrying vector
 CC plasmids incorporating the esterase secretory gene. The esterase
 CC is secreted from the host cell, and is recovered from the culture
 CC medium and from within the cells. The esterase is useful in
 CC hydrolysis reactions.
 SQ Sequence 588 AA;

Query Match 3.8%; Score 88; DB 19; Length 588;
 Best Local Similarity 23.3%; Pred. No. 4.10e+01;
 Matches 17; Conservative 19; Mismatches 33; Indels 4; Gaps 3;
 Db 1 vnfqfprneiadvirtrskvftvgiftafinllmlvpslymlqvdydrvlpsrneitilm 60
 Qy 176 IONYRCRGDDSKVQARKSFSGHASFMYT-MLYLV--LYLQARFTWR-GARLLRPLIQ 231
 Db 61 ltlmlgmfgmms 73
 Qy 232 FTLIMAFYTGLS 244

Search completed: Tue Nov 4 10:29:06 1997
 Job time : 78 secs.

This Page Blank (uspto)


```

ACCESSIONS S70114
REFERENCE S70114
#authors Fulton, L.
#submission submitted to the EMBL Data Library, May 1996
#description The sequence of S. cerevisiae cosmid 9819.
#accession S70114
#molecule_type DNA
##residues 1-289 #label FUL
##cross-references EMBL:U51031
GENETICS
#map_position 4R
SUMMARY #length 289 #molecular-weight 33514 #checksum 4703
Query Match 7.2%; Score 167; DB 11; Length 289;
Best Local Similarity 28.6%; Pred. No. 1.97e-09;
Matches 60; Conservative 46; Mismatches 91; Indels 13; Gaps 11;
Db 40 pferqfindltishyattervnmmlfyfsvpseltlllig--sl--ladrrhlifi 95
QY 61 PYHRGYCNDESIKYPLKTGETINDAVLCAGVIAIAIAITGEFYRIYLLKSRSTION 120
Db 96 lytsllglsia-w-fs-t-s-fftnfkwnigrlrpdldrcgp-vegpldtlftakdv 149
QY 121 PYAALYKQVCCFLGCATISQSTFDIAKVSIGRLRPHFLSVCPDFSQINCSGYQNTNR 180
Db 150 ctknherlldgfrtpsgshssesfaglyfwlcgqltlesplmplwrkmvafllpg 209
QY 181 CRG-DDSKVQEARKEFSFGHASFMTMLYLVLYLQARFTWRGRL-L-RPLQFTLIMM 237
Db 210 aalialartqdyrhfhvdvllgsmlyima 239
QY 238 AFTGLSRVSDHKHPDSVDLAGFAQAGALVA 267

RESULT 3
ENTRY S32217 #type complete
TITLE hypothetical protein 2 - Bacillus megaterium
ORGANISM #formal_name Bacillus megaterium
DATE 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change
ACCESSIONS S32217
REFERENCE S32215
#authors Rauschenbach, R.; Isernhagen, M.; Noeske-Jungblut, C.;
#submission submitted to the EMBL Data Library, March 1993
#description Cloning, sequencing and expression of the gene for cytochrome
p450meg, the steroid 15beta-monooxygenase from Bacillus
megaterium ATCC 13366.
#accession S32217
##status preliminary
#molecule_type DNA
##residues 1-216 #label RAU
##cross-references EMBL:221972
SUMMARY #length 216 #molecular-weight 24946 #checksum 8213
Query Match 5.9%; Score 138; DB 10; Length 216;
Best Local Similarity 38.6%; Pred. No. 2.46e-05;
Matches 32; Conservative 20; Mismatches 23; Indels 8; Gaps 5;
Db 127 sfpsghannafslgltflwrlthitarw--aril--llfslmli-lslglsriylgvh 181
QY 194 SFFSGHA--SFSMYTMLYLVLYLQARFTWRGRLRPLQFTLIMMAFYTLGRVSDHKH 251
Db 182 ypsdilatagylaggcwaisiwwf 204
QY 252 HPSDVLAFAGQA-LVACCIVFF 273

RESULT 4
ENTRY S58067 #type fragment
TITLE probable olfactory receptor tpcr34 - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 13-Jan-1996 #sequence_revision 19-Apr-1996 #text_change

```

```

ACCESSIONS S58067
REFERENCE S58067
#authors Vanderhaeghen, P.; Schurmans, S.; Vassart, G.; Parmentier, M.
#submission submitted to the EMBL Data Library, July 1995
#description Male germ cells from several mammalian species express a
specific repertoire of olfactory receptor genes.
#accession S58067
##status preliminary
#molecule_type mRNA
##residues_type 1-157 #label VAN
##cross-references EMBL:X89686
SUMMARY #length 157 #checksum 1240
Query Match 5.5%; Score 128; DB 14; Length 157;
Best Local Similarity 27.6%; Pred. No. 5.33e-04;
Matches 27; Conservative 30; Mismatches 33; Indels 8; Gaps 8;
Db 18 llvflsvlslfsvqlhnsvvlqlytkfsvdishfcdpsql-lnlacsdtdftnnlmvfy 76
QY 35 LLICLD-LFCLFMAGLP-FLIETSTIKPYHRG-FYCNDSEIKYPLKTGETI-NDAVLCA 90
Db 77 vgtisgfipi-sgiffs-yy-kivssilrmpspggyk 111
QY 91 VGVIVAILAITGEFYRIYLLKSRSTIONPYVAALYK 128

RESULT 5
ENTRY S28996 #type fragment
TITLE G protein-coupled receptor (clone PTE01) - rat (fragment)
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
ACCESSIONS S28996
REFERENCE S28996
#authors Abe, K.; Kusakabe, Y.; Tanemura, K.; Emori, Y.; Arai, S.
#journal FEBS Lett. (1993) 316:253-256
#title Multiple genes for G protein-coupled receptors and their
expression in lingual epithelia.
#accession S28996
#molecule_type mRNA
##residues 1-185 #label ABE
SUMMARY #length 185 #checksum 2049
Query Match 5.2%; Score 122; DB 14; Length 185;
Best Local Similarity 28.6%; Pred. No. 3.20e-03;
Matches 28; Conservative 30; Mismatches 32; Indels 8; Gaps 8;
Db 85 llvflsvlslfsvqlhnsvvlqlytkfsvdishfcdpsql-lnlacsdtdftnnlmvfy 143
QY 35 LLICLD-LFCLFMAGLP-FLIETSTIKPYHRG-FYCNDSEIKYPLKTGETI-NDAVLCA 90
Db 144 vgaigsfipi-sgiffs-yy-kivssilrmpspggyk 178
QY 91 VGVIVAILAITGEFYRIYLLKSRSTIONPYVAALYK 128

RESULT 6
ENTRY S50669 #type complete
TITLE hypothetical protein YER166w - yeast (Saccharomyces
cerevisiae)
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 28-May-1993 #sequence_revision 24-Feb-1995 #text_change
ACCESSIONS S50669; S30822
REFERENCE S50428
#authors Dietrich, F.S.
#submission submitted to the EMBL Data Library, December 1994
#description The sequence of S. cerevisiae cosmids 9163 and 9132.
#accession S50669
#molecule_type DNA
##residues 1-1571 #label DIE
##cross-references EMBL:U18922

```

```

REFERENCE S30812
#authors Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.;
      Komp, C.; Wei, Y.; Taylor, P.; Nakahara, K.; Roberts, D.;
      Davis, R.W.
#submission submitted to the EMBL Data Library, February 1993
#accession S30822
#molecule_type DNA
#residues 1-222 ##label MUL
##cross-references EMBL:L10718
GENETICS
#map_position 5R
SUMMARY #length 1571 #molecular-weight 177797 #checksum 2315
Query Match 5.2%; Score 120; DB 11; Length 1571;
Best Local Similarity 20.0%; Pred. No. 5.75e-03;
Matches 18; Conservative 28; Mismatches 40; Indels 3; Gaps 3;
Db 1244 vndtislvpqlyrvgilrkwnqrkflwylmldglyqslcifflylvyhkmvtsngl 1303
QY 94 VTAIAITGEFRIYYLKKSTIONP-YV-AALYKQVGCFLGCAISQSFTDIKYSI 151
Db 1304 gldhryfvgvvtiaviscntyvllhqyr 1333
QY 152 GRLRPHLSVCNPFDSQINCSE-GYIQNYR 180
RESULT 7
ENTRY S64327 #type complete
TITLE Probable membrane protein YGR036c - yeast (Saccharomyces
ALTERNATE_NAMES cerevisiae)
ORGANISM hypochetrical protein G4085
#formal_name Saccharomyces cerevisiae
DATE 17-May-1996 #sequence_revision 17-May-1996 #text_change
19-Jul-1996
ACCESSIONS S64327
REFERENCE S64071
#authors Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.
#submission submitted to the Protein Sequence Database, May 1996
#accession S64327
#molecule_type DNA
#residues 1-239 ##label RIE
##cross-references EMBL:Z72821
##experimental_source strain S288C
GENETICS
#map_position 7R
KEYWORDS transmembrane protein
FEATURE
39-55 #domain transmembrane #status predicted #label TM1\
136-152 #domain transmembrane #status predicted #label TM2\
169-185 #domain transmembrane #status predicted #label TM3\
SUMMARY #length 239 #molecular-weight 27649 #checksum 3184
Query Match 5.1%; Score 118; DB 12; Length 239;
Best Local Similarity 27.4%; Pred. No. 1.03e-02;
Matches 17; Conservative 21; Mismatches 22; Indels 2; Gaps 2;
Db 126 swknlfnlekciifsgalallsfcvfrvylhnlqdvigsvgtlgs-lyffivgi 184
QY 219 TWRGARLL-RPLQLFTLIMAWYTLGSRVSDHKHPSDVLGFAQALVACCVFVSDLI 277
Db 185 lr 186
QY 278 FK 279
RESULT 8
ENTRY E48909 #type fragment
TITLE G-protein coupled receptor Gpcr13 - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change
01-Dec-1995
ACCESSIONS E48909
REFERENCE A48909

```

```

#authors Wilkie, T.M.; Chen, Y.; Gilbert, D.J.; Moore, K.J.; Yu, L.;
      Simon, M.I.; Copeland, N.G.; Jenkins, N.A.
#journal Genomics (1993) 18:175-184
#title Identification, chromosomal location, and genome organization
      of mammalian G-protein-coupled receptors.
#accession E48909
#status preliminary; nucleic acid sequence not shown
#molecule_type mRNA
#residues 1-180 ##label WIL
##cross-references GB:L20334
SUMMARY #length 180 #checksum 3623
Query Match 4.8%; Score 111; DB 14; Length 180;
Best Local Similarity 37.8%; Pred. No. 7.57e-02;
Matches 17; Conservative 8; Mismatches 19; Indels 1; Gaps 1;
Db 130 vlcvtifsvllaivalyryifvvrsshadvagpqtlaalktv 174
QY 87 VLCAVGIVIAITLITGEFYRIYL-KKSRSTIQNPYYAALYKQV 130
RESULT 9
ENTRY JC1465 #type complete
TITLE Probable G protein-coupled receptor - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
13-Sep-1996
ACCESSIONS JC1465
REFERENCE JC1465
#authors Okazaki, H.; Ishizaka, N.; Sakurai, T.; Kurokawa, K.; Goto,
      K.; Kumada, M.; Takuwa, Y.
#journal Biochem. Biophys. Res. Commun. (1993) 190:1104-1109
#title Molecular cloning of a novel putative G protein-coupled
      receptor expressed in the cardiovascular system.
#accession JC1465
#molecule_type mRNA
#residues 1-352 ##label OKA
##experimental_source aortic smooth muscle
KEYWORDS glycoprotein; phosphoprotein; transmembrane protein
FEATURE
35-59 #domain transmembrane #status predicted #label TM1\
67-95 #domain transmembrane #status predicted #label TM2\
148-173 #domain transmembrane #status predicted #label TM3\
190-210 #domain transmembrane #status predicted #label TM4\
234-255 #domain transmembrane #status predicted #label TM5\
272-293 #domain transmembrane #status predicted #label TM6\
19 #binding_site carbohydrate (Asn) (covalent) #status
      predicted\
142,145,218,219, #binding_site phosphate (Ser) (covalent) #status
329,330,331,332 #binding_site phosphate (Thr) (covalent) #status
313 #binding_site phosphate (Thr) (covalent) #status
SUMMARY #length 352 #molecular-weight 38734 #checksum 3672
Query Match 4.8%; Score 111; DB 14; Length 352;
Best Local Similarity 37.8%; Pred. No. 7.57e-02;
Matches 17; Conservative 8; Mismatches 19; Indels 1; Gaps 1;
Db 191 vlcvtifsvllaivalyryifvvrsshadvagpqtlaalktv 235
QY 87 VLCAVGIVIAITLITGEFYRIYL-KKSRSTIQNPYYAALYKQV 130
RESULT 10
ENTRY S26033 #type complete
TITLE NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 -
      Caenorhabditis elegans mitochondrion (SGC4)
ORGANISM #formal_name mitochondrion Caenorhabditis elegans
DATE 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change
09-Sep-1994
ACCESSIONS S26033; S25806

```


Search completed: Tue Nov 4 10:27:30 1997
Job time : 82 secs.

This Page Blank (uspto)

RELEASE

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.

Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.

Distribution rights by IntelliGenetics, Inc.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Nov 6 11:36:00 1997; MasPar time 1135.34 Seconds
Tabular output not generated. 1317.542 Million cell updates/sec

Title: >US-08-842-827-5
Description: (1-1362) from US08842827.seq
Perfect Score: 1362
N.A. Sequence: 1 GGCGCAGCTCTGCAAAAGTT.....TTTTAAAAA.....AAAAA 1362
Comp: CCGCGTCGACGCTTTCAA.....AAATTTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 362067 seqs, 549138275 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb1-new3
1: BCT 2:FUN 3:GEN1 4:GEN2 5:HTG1 6:HTG2 7:HUM 8:INV
9:ORG 10:MAM 11:VRT 12:PLN 13:PRO 14:ROD 15:SYN 16:UNC
17:VIR

Database: genbank99
18:BCT1 19:BCT2 20:BCT3 21:BCT4 22:BCT5 23:BCT6 24:BCT7
25:BCT8 26:BCT9 27:BCT10 28:BCT11 29:GEN1 30:GEN2
31:GEN3 32:HTG1 33:HTG2 34:HTG3 35:INV1 36:INV2 37:INV3
38:INV4 39:INV5 40:INV6 41:INV7 42:INV8 43:INV9 44:INV10
45:INV11 46:MAM1 47:MAM2 48:MAM3 49:VRT1 50:VRT2 51:VRT3
52:VRT4 53:PAT1 54:PAT2 55:PAT3 56:PAT4 57:PAT5 58:PHG
59:PLN1 60:PLN2 61:PLN3 62:PLN4 63:PLN5 64:PLN6 65:PLN7
66:PLN8 67:PLN9 68:PLN10 69:PLN11 70:PRI1 71:PRI2
72:PRI3 73:PRI4 74:PRI5 75:PRI6 76:PRI7 77:PRI8 78:PRI9
79:PRI10 80:PRI11 81:PRI12 82:PRI13 83:PRI14 84:PRI15
85:ROD1 86:ROD2 87:ROD3 88:ROD4 89:ROD5 90:ROD6 91:ROD7
92:ROD8 93:ROD9 94:SYN 95:UNA 96:VRL1 97:VRL2 98:VRL3
99:VRL4 100:VRL5 101:VRL6 102:VRL7 103:VRL8 104:VRL9
105:VRL10
genbank-new3
106:BCT 107:GEN1 108:GEN2 109:HTG1 110:HTG2 111:INV
112:MAM 113:VRT 114:PHG 115:PLN 116:PRI1 117:PRI2
118:ROD 119:SYN 120:UNA 121:VRL
u-emb150_99
122:part1

Statistics: Mean 11.681; Variance 5.524; scale 2.114

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.	
1	1131	83.0	1444	78	HSU79294	Human clone 23748 mRNA	0.00e+00	
2	784	57.6	2206	91	RNDRI42	R.norvegicus mRNA for	0.00e+00	
3	102	7.5	1212	85	D84376	House mouse; Musculus	1.50e-56	
4	85	6.2	1622	40	DMU73822	Drosophila melanogast	8.36e-43	
5	71	5.2	1490	88	MUSHP1P	Mus musculus (clone H	8.69e-32	
c	6	35	215	57	128278	Sequence 5 from paten	6.42e-06	
7	33	2.4	215	57	128278	Sequence 5 from paten	1.12e-04	
c	8	28	2.1	1185	90	RATCEAC	Rat carcinoembryonic	9.77e-02
c	9	27	2.0	3203	79	HUMEMILX	Human prot-oncogene (3.50e-01
c	10	27	2.0	4204	91	RATNUCIA3	Rat nucleolin gene.	3.50e-01
c	11	26	1.9	4251	88	MUS45SRNA	Mouse 4.5S RNA gene.	1.22e+00
12	26	1.9	38118	33	CET7E9	*** SEQUENCING IN PRO	1.22e+00	
13	26	1.9	58382	110	CET28A8	Caenorhabditis elegans	1.22e+00	
14	26	1.9	58382	5	CET28A8	Caenorhabditis elegans	1.22e+00	
15	26	1.9	170789	110	CEY76A2	Caenorhabditis elegans	1.22e+00	
16	26	1.9	170789	5	CEY76A2	Caenorhabditis elegans	1.22e+00	
17	25	1.8	99	55	I13410	Sequence 12 from pate	4.09e+00	
18	24	1.8	201	53	A10161	Synthetic DNA for pre	1.33e+01	
c	19	24	1.8	201	53	A10159	S.griseus gene for pr	1.33e+01
c	20	25	1.8	801	50	GRSFR	G.gallus RSFR mRNA.	4.09e+00
c	21	24	1.8	870	85	MWGM2	M.musculus mRNA for h	1.33e+01
c	22	24	1.8	921	122	S78239	BSP-bone sialoprotein	1.33e+01
c	23	24	1.8	1400	13	RGU67155	Rubrivivax gelatinosu	1.33e+01
c	24	24	1.8	1770	38	CEU43316	Caenorhabditis elegans	1.33e+01
c	25	24	1.8	1914	69	YSPORF3	Schizosaccharomyces p	1.33e+01
c	26	24	1.8	2170	82	HUMOCT6A	Homo sapiens (oct-6)	1.33e+01
c	27	24	1.8	2502	122	HSB347	Human mRNA for Neurod	1.33e+01
c	28	24	1.8	2502	117	D82347	Human mRNA for Neurod	1.33e+01
c	29	24	1.8	2578	50	GGU67889	Gallus gallus bone si	1.33e+01
30	25	1.8	2708	41	DROZSTA	Drosophila virilis ze	4.09e+00	
31	24	1.8	2760	67	SCYOR382W	S.cerevisiae chromoso	1.33e+01	
32	24	1.8	2947	14	MWZIC	Mouse zic mRNA for zi	1.33e+01	
33	24	1.8	2947	90	MUSZIC	Mouse mRNA for zic pr	1.33e+01	
c	34	24	1.8	3595	77	HSU09825	Human acid finger pro	1.33e+01
35	24	1.8	3769	77	HSU11717	Human calcium activat	1.33e+01	
36	24	1.8	3987	77	HSU11058	Human calcium depende	1.33e+01	
37	24	1.8	4726	96	AVU48704	Adeno-associated viru	1.33e+01	
c	38	24	1.8	6789	94	ASPD3D3C	Artificial sequences	1.33e+01
39	24	1.8	7154	41	DROGYP5Y2	Drosophila virilis mo	1.33e+01	
40	24	1.8	22720	35	CEC34B7	Caenorhabditis elegans	1.33e+01	
41	25	1.8	34028	37	CELF59A6	Caenorhabditis elegans	4.09e+00	
42	25	1.8	37940	79	HUM19ANON1	Human DNA from cosmid	4.09e+00	
43	25	1.8	39320	36	CEFI3E6	Caenorhabditis elegans	4.09e+00	
44	24	1.8	184427	97	EHVU20824	Equine herpesvirus 2,	1.33e+01	
45	24	1.8	195158	34	HS473B4	Human DNA sequence **	1.33e+01	

ALIGNMENTS

RESULT 1 HSU79294 1444 bp mRNA PRI 30-JAN-1997
LOCUS Human clone 23748 mRNA, complete cds.
DEFINITION U79294
ACCESSION
NID g1710275

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 1444)

Andersson,B., Wentland,M.A., Ricafrente,J.Y., Liu,W. and Gibbs,R.A.

A 'double adaptor' method for improved shotgun library construction

Anal. Biochem. 236 (1), 107-113 (1996)

MEDLINE 96207227

REFERENCE 2 (bases 1 to 1444)

Yu,W., Andersson,B., Worley,K.C., Muzny,D.M., Ding,Y., Liu,W.,

Ricafrente,J.Y., Wentland,M.A., Lennon,G. and Gibbs,R.A.

Title Large Scale Concatenation cDNA Sequencing

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 1444)

[illegible]

SVCDPFSQINCEGYONYPGRGDSKVOEAKSFSGHASTSMETMLYLVLYLQAR
FTWRGAKLLPLLIQFTLLMAFYTGJSDRDKHPSDVLGAFQALVACCIVFFVS
DLEKFTKTLSPAPAREITLSPVDDINHHNMV"
503 a 634 c 583 g 486 t
BASE COUNT
ORIGIN

Query Match	57.6%	Score 784;	DB 91;	Length 2206;
Best Local Similarity	87.4%	Pred. No. 0.00e+00;		
Matches 951;	Conservative 0;	Mismatches 131;	Indels 6;	Gaps 4;

Db	396	ccgcagcagcgccatgcaagctacaagtagcagacaaggcgatcgctccctgagagtaaga	455
QY	280	ccgcagccagcccatgcanaactacaagtagcagacaaggcgatcgctccgggagagcaaga	339

Db 456 acggcgagcccgcgctcaacaacacccgaggaagggcgcagcaagcggtgctgc 515
|||||
QY 340 ACGGCGGACCCGGCGCTCAACAACCCGAGGGCGGCAAGCGGTGCTGC 399
|||||

[illegible]

Db 576 caagcaccaataagccttatcgctcgagggttttactgcaacgagagagcatcatcaagtacc 635
||||| ||||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QV 460 CAAGCACCATCAAGCCTTACCACCGAGGGTTTTACTGCAATGATGAGAGCATCAAGTACC 519

[illegible]

Db 696 ccattcctgggataatcacgggggaattctacgggatactattacctcaaggagagtc 755
|||||
Ov 580 ccattcctgggataatcacgggggaattctacgggatactattacctcaaggagagtc 635
|||||

Db 756 gatccaccattcagaacccttatgtggccgcctctctataagcaagtgggatgcttccttt 815

816 tcggctgtccatcagccagtccttcacagacatgcacaaagtgtccattggggcctga 875

Db 876 ggccctcacttcctcaagcgctctgtgacacctgatttcagtcagatcaattgctccagggct 935

936 acattcagaactacaggtgccgaggaggaagacagacagtagcaggaggccagggaatacct 995

Db 996 tttctcggggccacgctctcttccatgttccaaatgctgtatctggctgttatctctac 105

Db 1056 agccccgttcacctggcggtggggcccggtctctcgccccctctctgcagttcaattgc 111

Db 1116 tcattgatggccttctacacgggattgtcacgtgtatctgactacaacaccatctctagcg 117

557 TCATGATGGCCCTTCACACGGGACATGTCGCGGTATTCAGACCACAAAGCACCATCCCCAGTG 109
 Db 1176 atgtcctggcaggatttgcacaaggagctctggtggcctgctgcatagtagtgtcttcctgtg 123

QY	1057	ATGATCTGGCAGGATTGCTCAAGGAGGCCCCGGTGGGCCCTGCTGCATAGTATTTCTTCGCTGT	111
Db	1236	ctgacctcttcaagactaagagcacctctcactgcttgcctgccccgcgcatcaggagagaga	1299

QY	1117	CTGACCTCTCTCAAGACTAAGACGACGCTCTCCCTGCGCTGCTATCCGGAAGAAA	1117
Db	1296	tctgtctctctgtagacatcatggacaggaggaatcacacaacatggtgttaggtctgc	1355

QY 1177 TCC TTTCACCTGTGGACATTATTGACAGGAACAATCACCACAACATGATGATAGGTGCCAC 123

Db	1356	gg -cctccagacgactttctctgaagcgactgttgcacagaatgccctgtgtctctcca	1414
Qy	1237	CCACCTCCTGAGCTGTTTTGTAAATGACTGCT -GACAGCAAGTTCTTGCTGCTCTCCA	1295
Db	1415	atctcatcagcacagtagaatatagggaagaagcttttttcgcgatttgattttaaaacat	1474
Qy	1296	ATTCATCATCAGACAGTAGAATGTAGGGAAAACTTTTT -GCCCGCTGATTTTTTAAAAAAA	1354
Db	1475	ttaaagaa	1482
Qy	1355	AAAAAAA	1362

RESULT	3	D84376	1212 bp	mRNA	ROD	03-OCT-1996
LOCUS						
DEFINITION				House mouse; Musculus domesticus kidney mRNA for Phosphatidic acid phosphatase, complete cds.		

D84376
 ACCESSION
 NID
 91487872
 KEYWORDS
 phosphatidic acid phosphatase; 35-kDa phosphatidic acid phosphatase.

SOURCE	ORGANISM
Mus musculus kidney cDNA to mRNA.	
Mus musculus	mitochondrial eukaryotes; Metazoa; Chordata;
Mus musculus	Eukaryotae;
	Vertebrata;
	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae.

Mus.
1 (bases 1 to 1212)
Kano, H.
Direct Submission
REFERENCE
AUTHORS
TITLE

JOURNAL
Submitted (10-APR-1996) to the DDBJ/EMBL/GenBank databases. Video
Kano, Sapporo Medical University School of Medicine, Department of
Biochemistry, West-17, South-1, Sapporo, Hokkaido 060, Japan
4501-4502, 4503, 4504, 4505, 4506, 4507, 4508, 4509, 4510, 4511, 4512, 4513, 4514, 4515, 4516, 4517, 4518, 4519, 4520, 4521, 4522, 4523, 4524, 4525, 4526, 4527, 4528, 4529, 4530, 4531, 4532, 4533, 4534, 4535, 4536, 4537, 4538, 4539, 4540, 4541, 4542, 4543, 4544, 4545, 4546, 4547, 4548, 4549, 4550, 4551, 4552, 4553, 4554, 4555, 4556, 4557, 4558, 4559, 4560, 4561, 4562, 4563, 4564, 4565, 4566, 4567, 4568, 4569, 4570, 4571, 4572, 4573, 4574, 4575, 4576, 4577, 4578, 4579, 4580, 4581, 4582, 4583, 4584, 4585, 4586, 4587, 4588, 4589, 4590, 4591, 4592, 4593, 4594, 4595, 4596, 4597, 4598, 4599, 4600, 4601, 4602, 4603, 4604, 4605, 4606, 4607, 4608, 4609, 4610, 4611, 4612, 4613, 4614, 4615, 4616, 4617, 4618, 4619, 4620, 4621, 4622, 4623, 4624, 4625, 4626, 4627, 4628, 4629, 4630, 4631, 4632, 4633, 4634, 4635, 4636, 4637, 4638, 4639, 4640, 4641, 4642, 4643, 4644, 4645, 4646, 4647, 4648, 4649, 4650, 4651, 4652, 4653, 4654, 4655, 4656, 4657, 4658, 4659, 4660, 4661, 4662, 4663, 4664, 4665, 4666, 4667, 4668, 4669, 4670, 4671, 4672, 4673, 4674, 4675, 4676, 4677, 4678, 4679, 4680, 4681, 4682, 4683, 4684, 4685, 4686, 4687, 4688, 4689, 4690, 4691, 4692, 4693, 4694, 4695, 4696, 4697, 4698, 4699, 4700, 4701, 4702, 4703, 4704, 4705, 4706, 4707, 4708, 4709, 4710, 4711, 4712, 4713, 4714, 4715, 4716, 4717, 4718, 4719, 4720, 4721, 4722, 4723, 4724, 4725, 4726, 4727, 4728, 4729, 4730, 4731, 4732, 4733, 4734, 4735, 4736, 4737, 4738, 4739, 4740, 4741, 4742, 4743, 4744, 4745, 4746, 4747, 4748, 4749, 4750, 4751, 4752, 4753, 4754, 4755, 4756, 4757, 4758, 4759, 4760, 4761, 4762, 4763, 4764, 4765, 4766, 4767, 4768, 4769, 4770, 4771, 4772, 4773, 4774, 4775, 4776, 4777, 4778, 4779, 4780, 4781, 4782, 4783, 4784, 4785, 4786, 4787, 4788, 4789, 4790, 4791, 4792, 4793, 4794, 4795, 4796, 4797, 4798, 4799, 4800, 4801, 4802, 4803, 4804, 4805, 4806, 4807, 4808, 4809, 4810, 4811, 4812, 4813, 4814, 4815, 4816, 4817, 4818, 4819, 4820, 4821, 4822, 4823, 4824, 4825, 4826, 4827, 4828, 4829, 4830, 4831, 4832, 4833, 4834, 4835, 4836, 4837, 4838, 4839, 4840, 4841, 4842, 4843, 4844, 4845, 4846, 4847, 4848, 4849, 4850, 4851, 4852, 4853, 4854, 4855, 4856, 4857, 4858, 4859, 4860, 4861, 4862, 4863, 4864, 4865, 4866, 4867, 4868, 4869, 4870, 4871, 4872, 4873, 4874, 4875, 4876, 4877, 4878, 4879, 4880, 4881, 4882, 4883, 4884, 4885, 4886, 4887, 4888, 4889, 4890, 4891, 4892, 4893, 4894, 4895, 4896, 4897, 4898, 4899, 4900, 4901, 4902, 4903, 4904, 4905, 4906, 4907, 4908, 4909, 4910, 4911, 4912, 4913, 4914, 4915, 4916, 4917, 4918, 4919, 4920, 4921, 4922, 4923, 4924, 4925, 4926, 4927, 4928, 4929, 4930, 4931, 4932, 4933, 4934, 4935, 4936, 4937, 4938, 4939, 4940, 4941, 4942, 4943, 4944, 4945, 4946, 4947, 4948, 4949, 4950, 4951, 4952, 4953, 4954, 4955, 4956, 4957, 4958, 4959, 4960, 4961, 4962, 4963, 4964, 4965, 4966, 4967, 4968, 4969, 4970, 4971, 4972, 4973, 4974, 4975, 4976, 4977, 4978, 4979, 4980, 4981, 4982, 4983, 4984, 4985, 4986, 4987, 4988, 4989, 4990, 4991, 4992, 4993, 4994, 4995, 4996, 4997, 4998, 4999, 5000, 5001, 5002, 5003, 5004, 5005, 5006, 5007, 5008, 5009, 5010, 5011, 5012, 5013, 5014, 5015, 5016, 5017, 5018, 5019, 5020, 5021, 5022, 5023, 5024, 5025, 5026, 5027, 5028, 5029, 5030, 5031, 5032, 5033, 5034, 5035, 5036, 5037, 5038, 5039, 5040, 5041, 5042, 5043, 5044, 5045, 5046, 5047, 5048, 5049, 5050, 5051, 5052, 5053, 5054, 5055, 5056, 5057, 5058, 5059, 5060, 5061, 5062, 5063, 5064, 5065, 5066, 5067, 5068, 5069, 5070, 5071, 5072, 5073, 5074, 5075, 5076, 5077, 5078, 5079, 5080, 5081, 5082, 5083, 5084, 5085, 5086, 5087, 5088, 5089, 5090, 5091, 5092, 5093, 5094, 5095, 5096, 5097, 5098, 5099, 5100, 5101, 5102, 5103, 5104, 5105, 5106, 5107, 5108, 5109, 5110, 5111, 5112, 5113, 5114, 5115, 5116, 5117, 5118, 5119, 5120, 5121, 5122, 5123, 5124, 5125, 5126, 5127, 5128, 5129, 5130, 5131, 5132, 5133, 5134, 5135, 5136, 5137, 5138, 5139, 5140, 5141, 5142, 5143, 5144, 5145, 5146, 5147, 5148, 5149, 5150, 5151, 5152, 5153, 5154, 5155, 5156, 5157, 5158, 5159, 5160, 5161, 5162, 5163, 5164, 5165, 5166, 5167, 5168, 5169,

REFERENCE

2 (bases 1 to 1212)

Kai, M., Wada, I., Imai, S., Sakane, F., and Kanoh, H.

AUTHORS

phosphatase (type 2) bound to plasma membranes. Polymerase chain reaction amplification of mouse H2O2-inducible hlc53 clone yielded the cDNA encoding phosphatidic acid phosphatase

96324980	Location/Qualifiers
FEATURES	1..1212
MEDLINE	source

```

cnds
/organism="Mus musculus"
/tissue_type="kidney"
287..1138
evidence=experimental

```

```

/condition=stat-1
/product="phosphatidic acid phosphatase"
/db_xref="PID:G1487873"
/translation="MEDKRLPYVALDVICLLAGLFFAILTSRTPFPQFGICNDSDS

```

[illegible]

BASE COUNT	ISINHEP"			
277 a	316 c	313 g	306 t	
ORIGIN				

```

Query Match          7.5%;      Score 109;      DB 85;      Length 1212;
Best Local Similarity 62.1%;      Pred. NO;      1.50e-56;
Matches 261;      Conservative 0;      Mismatches 159;      Indels 0;      Gaps 0;

```

DB 620 agtcagctcttgaactgcactgaatataactatagcagatttgcggcgcaactcttg 679
QY 711 AGCCAGTCTTTACAGAGCATTGCCAAAGTGTCCATAGGGCGCCTGCCTCTCACTCTTG 770

Db 680 gctatctgaacccagactggtcaaaatcaactgcagtgatggctattatgaggactac 739
2Y 771 AGTGTCTGCAACCTGATTTTCACCCAGATCACTGCTCTGAAGGCTCATTCAGAACTAC 830

```

Db 740 atatgtcaagggaatgaagaagaaagcaaggaggggaggtgtctcttctactctgggacac 799
QY 831 AGATGAGAGGTGATGACAGCAAGTCCAGGAAGCAGGAGTCCCTTCTCTCTGCGCCAT 890
Db 800 tcttctctctactgactgactggtgtgtgcgactttatcttcaagccaggaagaag 859
QY 891 GCCTCTCTCCATGTACACTATGCTGTATTTGGTGCTATACCTGCAGGCCGCTTCACT 950
Db 860 ggagactgggcaagactcttaccagcccatgctccagcttgggtgcactgtcttttccata 919
QY 951 TGCGAGGAGCGCGCTGCTCGCGCCCTCCCTGCATCTACCTTGATGATGCGCTTC 1010
Db 920 tatgtggccttctctgagtgtgtgactacaacacacacagtgagtgactcaagttgga 979
QY 1011 TACAGGGAGTGTCTCGCTGCTCAGACCAACAGCAGCATCCAGCTATGCTTGGCAGGA 1070
Db 980 ctcttcagggagctctgatgctatactggtgtcttcttatgtatcccatcttccaag 1039
QY 1071 TTGCTCAAGGAGCCCTGCTGCGCTGCTGCATAGTTTCTTCTGCTGCTGACCTCTTCAAG 1130

RESULT 4
LOCUS DMU73822 1622 bp DNA INV 07-FEB-1997
DEFINITION Drosophila melanogaster wunen gene, complete cds.
ACCESSION U73822
NID 91769524
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
REFERENCE 1 (bases 1 to 1622)
AUTHORS Zhang, N., Zhang, J.-P., Purcell, K., Chen, Y. and Howard, K.
TITLE The Drosophila protein Wunen repels migrating germ cells
JOURNAL Nature 385 (5611), 64-67 (1997)
MEDLINE 97138220
REFERENCE 2 (bases 1 to 1622)
AUTHORS Zhang, N., Zhang, J.-P., Purcell, K., Chen, Y. and Howard, K.
TITLE Direct Submission
JOURNAL Submitted (10-OCT-1996) Molecular Cell Biology, University College, Gower, London WC1E 6BT, England
FEATURES
source
location/Qualifiers
1..1622
/organism="Drosophila melanogaster"
/chromosome="2"
/map="45D"
197..1099
/feature="germ cell guidance factor; similar to phosphatidic acid phosphatase type 2"
/codon_start=1
/db_xref="pid:g1769525"
/translation="MDTKRLRVLGVLDVLLLCAGFPILLFLGLGEPYKRGFFCDE
SLKHPFDSTWMLYIGAVIPGVILIVEIISQNKAKDQNGNATSRXYFMYE
LPDMWIECKTIGYAFAGLSQLTDAKISIGRLRPHFTAVCPQADSGTCDIAI
NAGYIQBCTCKGSSARMKEMRLSPSGHSFTFFAMVYLYLQARMTWRGSKL
LRHLQFLFIMVWYATLSRVSVDYKHSVDVLAGSLGISALVAVNVSDLFXKPNT
KEYLARTVQDMNASPAQITITN"
BASE COUNT 469 a 392 c 367 g 383 t 11 others
ORIGIN
Query Match 6.2%; Score 85; DB 40; Length 1622;
Best Local Similarity 71.4%; Pred. No. 8.36e-43;
Matches 142; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Db 755 tcttcccaagtggaactccagcttcaacttcttgcacatggtctatctggcgctctac 814
QY 873 TCCTTCTCTGCGCATGCTCTCTCCATGCTATACATGCTGTATTTGGTGCTATAC 932
Db 815 ctgcagctcgatgactgaggggctcaagctactgctgacaccttctccagttctcg 874
QY 933 CTGCAGGCCCGCTTCACTTGGCGAGGAGCGCGCTGTCCGCGCCCTCTCTGCAAGTTCACC 992

```

```

Db 875 ttcatcatggtgctgtgtacacagccctaaagtcgctatcgactacaagcaccactgg 934
QY 993 TTGATCATGATGCGCTTCTACACGGGACTGTCTCGGTATACAGACCAAGCACCATCCC 1052
Db 935 tccgatgtgctgcsaggat 953
QY 1053 AGTGATGTTCTGCGCAGGAT 1071

RESULT 5
LOCUS MUSHPIP 1490 bp mRNA ROD 11-JAN-1996
DEFINITION Mus musculus (clone HIC-53) hydrogen peroxide-inducible protein mRNA, complete cds.
ACCESSION L43371
NID g1161099
KEYWORDS hydrogen peroxide; hydrogen peroxide-inducible protein.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1490)
AUTHORS Egawa, K., Yoshiwara, M., Shibamura, M. and Nose, K.
TITLE Isolation of a novel ras-recision gene that is induced by hydrogen peroxide from a mouse osteoblastic cell line, MC3T3-E1
JOURNAL FEBS Lett. 372 (1), 74-77 (1995)
MEDLINE 96032549
FEATURES
source
location/Qualifiers
1..1490
/organism="Mus musculus"
/clone="HIC-53"
/cell_line="MC3T3-E1"
/dev_stage="embryo"
/tissue_type="calvariagag"
<1..978
/codon_start=1
/product="hydrogen peroxide-inducible protein"
/db_xref="pid:g1161100"
/translation="ARASLELARVPVGGTGORRPIGRDEPALRQAEPAQIAASTVVK
ESRKHVSRRPIOTGPGPCSRQKRLAVANGRRHPEGALGIGYLDRRGLPLPLAP
GGDTIQPTVMFDKTRLPYVALDVLCVLAGLPFALTSRHTPFGRGIFCNDSDSIRPY
KEDTIPYALGLGIVIPFCIIIVMSIGESLSVFNVLHNSFVNPVIATIIYKAVGALF
GVSASQSLTDIAKVTIGSLRPHLAICNPDMSKINCSDGYIEDYICQNEEKVKRGRL
SFGCTLRLTYLVHAVCRTLSSRLRLRLRPLQFLIAFSIYVGLFSV"
<1..1490
BASE COUNT 362 a 367 c 368 g 393 t
ORIGIN
Query Match 5.2%; Score 71; DB 88; Length 1490;
Best Local Similarity 67.8%; Pred. No. 8.69e-32;
Matches 154; Conservative 0; Mismatches 71; Indels 2; Gaps 2;

Db 667 agtcagctcttgaactgacatgctaatatatactatcagcagcttgcgcccagctcttg 726
QY 711 ACCGAGTCTTTTCAGACATATCCCAAGTGTCCATAGGCGCGCTCGCTCCTCTCTTG 770
Db 727 gctatctgtacccagactggtcctaaatcaactgagtgatgcttatattgagactac 786
QY 771 AGTGCTGCAACCCCTGATTTTCAGCCAGATCAACTGCTCTGAAGGCTACATTGAGAACTAC 830
Db 787 atattgcaagggaatgaagaagaactcaaggagggcaggtgtgtttct-ctg-ggacac 844
QY 831 AGATCAGAGGTGATGACAGCAAGTCCAGGAAGCCAGGAAGTCTCTCTCTCTGCGCAT 890
Db 845 tcttctctctctgactgctgctgtttgttcgcactttatcttca 891
QY 891 GCCTCTCTCTCCTCATGCTATGCTATTTGGTGCTATACCTGCA 937

RESULT 6

```

Handwriting practice lines for the letter 'L' in various styles and sizes.

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

```
MParch_pp  protein - protein database search, using Smith-Waterman algorithm
Run on:      Tue Nov  4 10:25:03 1997;  MasPar time 8.34 Seconds
Tabular output not generated.  791.390 Million cell updates/sec
```

```
>US-08-842-827-6
Description: (1-311) from US08842827.pep
Perfect Score: 2326
Sequence: 1 MQNYKYDKAIVPSKNGSP.....RKEILSPVDIIRNNHHNM 311
```

Scoring table: PAM 150
Gap 11

Searched: 59021 seqs, 21210388 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

```
Database:      swiss-prot34
               1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
               8:part8 9:part9 10:part10 11:part11
```

Statistics: Mean 47.573; Variance 83.956; scale 0.567

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %			Length	DB	ID	Description	Pred. No.
	Score	Match	Length					
1	431	18.5	341	11	YX3_CAEEL		HYPOTHETICAL 39.0 KD	1.55e-66
2	122	5.2	185	4	G01_YAT		POSSIBLE CUSTODIUM RE	1.43e-04
3	120	5.2	171	1	ATC5_YEAST		PROBABLE CALCIUM-TRAN	2.91e-04
4	118	5.1	239	11	YG1_YEAST		HYPOTHETICAL 27.6 KD	5.91e-04
5	111	4.8	180	3	EDGL_MOUSE		PROBABLE G PROTEIN-CO	6.63e-03
6	111	4.8	352	4	H218_RAT		PROBABLE G PROTEIN-CO	6.63e-03
7	112	4.8	409	7	NU4M_CAEEL		NADH-UBIQUINONE OXID	4.72e-03
8	103	4.4	434	11	YEAR_BACSU		HYPOTHETICAL 46.4 KD	9.31e-02
9	100	4.3	320	10	VU25_HSVJ7		U25 PROTEIN.	2.42e-01
10	101	4.3	346	11	YI3_YEAST		HYPOTHETICAL 37.4 KD	1.76e-01
11	99	4.3	354	11	YVCU_YEAS		YOP PROTEINS TRANSLOC	3.31e-01
12	101	4.3	404	6	LPM1_EBVC		LATENT MEMBRANE PROTE	1.76e-01
13	101	4.3	529	2	CIK6_HUMAN		POTASSIUM CHANNEL PRO	1.76e-01
14	97	4.2	208	10	YAG1_LEGPN		HYPOTHETICAL 23.7 KD	6.15e-01
15	97	4.2	332	9	SRG3_CAEEL		SRG-3 PROTEIN.	6.15e-01
16	97	4.2	386	6	LPM1_EBV		LATENT MEMBRANE PROTE	6.15e-01
17	97	4.2	386	6	LPM1_EBVR		LATENT MEMBRANE PROTE	6.15e-01
18	98	4.2	614	6	NTBE_CANFA		SODIUM- AND CHLORIDE-K	4.52e-01
19	97	4.2	2493	11	YPM4_YEAST		HYPOTHETICAL 287.5 KD	6.15e-01
20	98	4.2	3898	7	POLG_BVDVS		GENOME POLYPROTEIN.	4.52e-01
21	96	4.1	253	6	LEP3_VIBCH		TYPE 4 PREPILIN-LIKE	8.35e-01
22	95	4.1	263	6	MIP_RANPI		LENS FIBER MAJOR INTR	1.13e+00

RA HYMAN R., KAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.,

AC P52592;

AC P52592;

```

FT TRANSMEM      35   59    1 (POTENTIAL).
FT DOMAIN        60   66    CYTOPLASMIC (POTENTIAL).
FT TRANSMEM     67   95    2 (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
FT DOMAIN       96  109    3 (POTENTIAL).
TRANSMEM      110  128    CYTOPLASMIC (POTENTIAL).
FT DOMAIN      129  147    4 (POTENTIAL).
TRANSMEM      148  173    EXTRACELLULAR (POTENTIAL).
FT DOMAIN      174  189    5 (POTENTIAL).
TRANSMEM      190  210    CYTOPLASMIC (POTENTIAL).
FT DOMAIN      211  233    6 (POTENTIAL).
TRANSMEM      234  255    EXTRACELLULAR (POTENTIAL).
FT DOMAIN      256  271    7 (POTENTIAL).
TRANSMEM      272  292    CYTOPLASMIC (POTENTIAL).
FT DOMAIN      293  352    POTENTIAL.
FT CARBOHYD     19   19    PALMITATE (BY SIMILARITY).
FT LIPID        305  305    PALMITATE; 2F532027 CRC32;
SQ SEQUENCE    352 AA; 38734 MW; 2F532027 CRC32;

Query Match          4.8%; Score 111; DB 4; Length 352;
Best Local Similarity 37.8%; Pred.No. 6.63e-03;
Matches 17; Conservative 8; Mismatches 19; Indels 1; Gaps

Db 191 vlcvltifslaiavalyriyfvvrsshadvagptlalkktv 235
||| ||| : | : |:: : |: : || :: || |
QY 87 VLCAGVIVAILIIGEFRIYL-KKSRSTQNPFVAALYKQV 130

RESULT           7
ID NU4M_CAEEL STANDARD; PRT; 409 AA.
AC P24892;
DT 01-MAR-1992 (REL. 21, CREATED)
DI 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DE 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE NADH-UBIUINONE OXIDOREDUCTASE CHAIN 4 (EC 1.6.5.3).
GN ND4.
OS CAENORHABDITIS ELEGANS.
OG MITOCHONDRION.
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIIDA.
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 92201635.
RA OKIMOTO R., MACFARLANE J.L., CLARY D.O., WOLSTENHOLME D.R.;
RL GENETICS 130:471-498(1992).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CD ENBL; X54252; G559502; -.
DR PIR; S26033; S26033.
RS OXIDOREDUCTASE; NAD; UBIQUINONE; MITOCHONDRIUM.
SK SEQUENCE 409 AA; 47206 MW; DC040448 CRC32;

Query Match          4.8%; Score 112; DB 7; Length 409;
Best Local Similarity 27.6%; Pred.No. 4.72e-03;
Matches 21; Conservative 25; Mismatches 22; Indels 8; Gaps

Db 18 flflftvmfslfnnn-fswgglflvidysfililivmslfilgilvis-knn--nlli- 72
| ::::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 202 FSNATMLVLVLQLARETWGARGALLRLLQLLFT-LTLMAYVT-GLSRVSDKHHPDVLAG 259
| : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Db 73 lse-ilvficiliiffip 87
::: || ||::|::

QY 260 FAQGALVACCIVEFFVS 275

RESULT           8
ID YBAR_BACSU STANDARD; PRT; 434 AA.
AC AC P55189;
DT 01-OCT-1996 (REL. 34, CREATED)
DI 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DE DE HYPOTHETICAL 46.4 KD PROTEIN IN RRNG-FEUC INTERGENIC REGION.
GN YEAR.
OS BACILLUS SUBTILIS.
OC PROKARYOTA; FIRMICUTES; ENDOSPHERE-FORMING RODS AND COCCI; BACILLACEA
```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RA LIU H., YASUMOTO K., HAGA K., OHASHI Y., YOSHIKAWA H., TAKAHASHI H.;
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -|- SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.
DR EMBL; D84213; G1256150; -.
KW SUBTILIST; BG11562; YBAR.
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE; TRANSPORT.
FT TRANSMEM 4 24 POTENTIAL.
FT TRANSMEM 43 63 POTENTIAL.
FT TRANSMEM 77 97 POTENTIAL.
FT TRANSMEM 99 119 POTENTIAL.
FT TRANSMEM 124 144 POTENTIAL.
FT TRANSMEM 176 196 POTENTIAL.
FT TRANSMEM 215 235 POTENTIAL.
FT TRANSMEM 251 271 POTENTIAL.
FT TRANSMEM 273 293 POTENTIAL.
FT TRANSMEM 301 321 POTENTIAL.
FT TRANSMEM 322 342 POTENTIAL.
SQ SEQUENCE 434 AA; 46435 MW; D5B8B2CD CRC32;

Query Match 4.4%; Score 103; DB 11; Length 434;
Best Local Similarity 37.5%; Pred. No. 9,31e-02;
Matches 21; Conservative 13; Mismatches 18; Indels 4; Gaps 3;

Db 97 feg-aswsmvamlagslvilvlprrfttapsplvaimvmtliavtfhdvrtvgd 151
QY 196 FSHGASFMYTML--YLVL-YLQARTWRGARLLRPLLOFTLIMMAFYTLGLSRVSD 248

RESULT 9
ID VU25_HSV7J STANDARD; PRT; 320 AA.
AC P52528;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE U25 PROTEIN.
GN U25.
OS HERPES SIMPLEX VIRUS (TYPE 7 / STRAIN JI) (HHV7).
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; BETAHERPESVIRINAE.
RN [1]
RA NICHOLAS J.;
RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -|- SIMILARITY: BELONGS TO US22 FAMILY. STRONGEST SIMILARITY TO
CC HUMAN CYTOMEGALOVIRUS UL43 PROTEIN.
DR EMBL; U43400; G1139627; -.
SQ SEQUENCE 320 AA; 37883 MW; 81467893 CRC32;

Query Match 4.3%; Score 100; DB 10; Length 320;
Best Local Similarity 20.7%; Pred. No. 2.42e-01;
Matches 29; Conservative 42; Mismatches 60; Indels 9; Gaps 9;

Db 44 etipeisksdieirntylc-cketiivgvylh-hvkkstcrgp-lllgdgrhlyvng 100
QY 141 QSTFDIAKVSIGRLRHFSLVSCNPFDSQINCSEGYIQNYRCRGDDSKVQEARSKFFSGHA 200

Db 101 ffdk-alyvssnlqdfll-vglkffpyiyeledfivdfetg-skivehsksfsmli-y 156
QY 201 SFSMYTMLVLVLVYLQARTWRGARLLRPLLOFTLIMMAFYTLGLSRVSDHKKHPSDVLAGE 260

Db 157 rden-invfcil-ksspykt 174
QY 261 AQGALVACCIVFFVSDLFKT 280

RESULT 10
ID YIC3_YEAST STANDARD; PRT; 346 AA.
AC P40544;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)

```

```

DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 37.4 KD PROTEIN IN SSM4-MP11 INTERGENIC REGION.
GN YIL023C.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-5288C / AB972;
RA BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,
RA CHURCHER C.M., CONNOR R., COPSEY T., DEAR S., DEVLIN K., FRASER A.,
RA GENTLES S., HAMLYN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,
RA LOUIS E., LYE G., MOULE S., MOULE T., ODELL C., PEARSON D.,
RA RAJANDREAM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,
RA WALSH S.V., WHITEHEAD S.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
DR EMBL; 246881; G599976; -.
DR EMBL; 247047; G763323; -.
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 70 90 POTENTIAL.
FT TRANSMEM 100 120 POTENTIAL.
FT TRANSMEM 127 147 POTENTIAL.
FT TRANSMEM 203 223 POTENTIAL.
FT TRANSMEM 253 273 POTENTIAL.
FT TRANSMEM 291 311 POTENTIAL.
FT TRANSMEM 323 343 POTENTIAL.
SQ SEQUENCE 346 AA; 37418 MW; DE70B3A8 CRC32;

Query Match 4.3%; Score 101; DB 11; Length 346;
Best Local Similarity 21.7%; Pred. No. 1.76e-01;
Matches 26; Conservative 41; Mismatches 45; Indels 8; Gaps 8;

Db 22 vhnrdhghean-hesksqflllkqesifyslvvcflqnhflvlgprynaivailliqImp 80
QY 176 IQNYRCRGDDSKVQEARSKFFSGHAFSMYTML-VL-V-LY-LQARTWRGARLLRPLQ 231

Db 81 -clfv1-fvpgl-rkndrasltslsvsfslgtllgdlllhvipeslsgvtadvnvggsi 137
QY 232 FTLIMMAFYTLGLSRVSDHKKHPSDVLAGEAQAALVACCIVFFVSDLFKTTLTSLPAPAI 291

RESULT 11
ID YSCU_YERPS STANDARD; PRT; 354 AA.
AC P40300;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE YOP PROTEINS TRANSLLOCATION PROTEIN U.
GN YSCU.
OS YERSINIA PSEUDOTUBERCULOSIS.
OC PLASMID PIB1.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YPIII;
RX MEDLINE; 94222840.
RA BERGMAN T., ERICKSON K., GALYOV E., PERSSON C., WOLF-WATZ H.;
RL J. BACTERIOL. 176:2619-2626(1994).
CC -|- FUNCTION: COMPONENT OF THE YOP SECRETION MACHINERY.
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -|- SIMILARITY: BELONGS TO THE FLHB/HRPN/YSCU/SPAS FAMILY.
DR EMBL; L25667; G475126; -.
KW PLASMID; VIRULENCE; TRANSMEMBRANE.
FT TRANSMEM 30 50 POTENTIAL.
FT TRANSMEM 79 99 POTENTIAL.
FT TRANSMEM 138 158 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
FT TRANSMEM 187 207 POTENTIAL.
SQ SEQUENCE 354 AA; 40381 MW; AC150C3F CRC32;

Query Match 4.3%; Score 99; DB 11; Length 354;

```

WPALEA

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Thu Nov 6 12:23:32 1997; MasPar time 1029.95 Seconds
Tabular output not generated. 1313.733 Million cell updates/sec

Title: >US-08-842-827-7
Description: (1-1232) from US08842827.seq
Perfect Score: 1232
N.A. Sequence: 1 ACCATGCGAGGAGTGGGT.....CAAAAAAAAAAAAAAAAAA 1232
Comp: TGGTACGTCGCTCCACCA.....GTTTTTTTTTTTTTTTTT

Scoring table: TABLE default
Gap 6
Nmatch STD : Dbase 0: Query 0
Searched: 362067 seqs, 549138275 bases x 2
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: emb1-new3
1: BCT 2: FUN 3: GEN1 4: GEN2 5: HTG1 6: HTG2 7: HUM 8: INV
9: ORG 10: MAM 11: VRT 12: PLN 13: PRO 14: ROD 15: SYN 16: UNC
17: VIR
genbank99
18: BCT1 19: BCT2 20: BCT3 21: BCT4 22: BCT5 23: BCT6 24: BCT7
25: BCT8 26: BCT9 27: BCT10 28: BCT11 29: GEN1 30: GEN2
31: GEN3 32: HTG1 33: HTG2 34: HTG3 35: INV1 36: INV2 37: INV3
38: INV4 39: INV5 40: INV6 41: INV7 42: INV8 43: INV9 44: INV10
45: INV11 46: MAM1 47: MAM2 48: MAM3 49: VRT1 50: VRT2 51: VRT3
52: VRT4 53: PAT1 54: PAT2 55: PAT3 56: PAT4 57: PAT5 58: PHG
59: PLN1 60: PLN2 61: PLN3 62: PLN4 63: PLN5 64: PLN6 65: PLN7
66: PLN8 67: PLN9 68: PLN10 69: PLN11 70: PRI1 71: PRI2
72: PRI3 73: PRI4 74: PRI5 75: PRI6 76: PRI7 77: PRI8 78: PRI9
79: PRI10 80: PRI11 81: PRI12 82: PRI13 83: PRI14 84: PRI15
85: ROD1 86: ROD2 87: ROD3 88: ROD4 89: ROD5 90: ROD6 91: ROD7
92: ROD8 93: STR 94: SYN 95: UNA 96: VRL1 97: VRL2 98: VRL3
99: VRL4 100: VRL5 101: VRL6 102: VRL7 103: VRL8 104: VRL9
105: VRL10
genbank-new3
106: BCT 107: GEN1 108: GEN2 109: HTG1 110: HTG2 111: INV
112: MAM 113: VRT 114: PHG 115: PLN 116: PRI1 117: PRI2
118: ROD 119: SYN 120: UNA 121: VRL
Database: u-emb150_99
122: part1

Statistics: Mean 11.574; Variance 6.325; scale 1.830

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	127	10.3	1212 85	D84376	House mouse; Musculus	1.26e+66
2	119	9.7	2206 91	RNDRI42	R.norvegicus MRNA for	7.79e-61
3	106	8.6	1444 78	HSU79294	Human clone 23748 mRN	1.84e-51
4	59	4.8	1490 88	MUSHPIP	Mus musculus (clone H	2.95e-19
5	37	3.0	1622 40	DMU73822	Drosophila melanogast	8.42e-06
6	36	2.9	215 57	128278	Sequence 5 from paten	3.05e-05
7	33	2.7	215 57	128278	Sequence 5 from paten	1.33e-03
8	29	2.4	1104 80	HUMCOOTAA	Human CO-029.	1.58e-01
9	28	2.3	1943 92	S42292	mannose-binding prote	4.94e-01
10	28	2.3	1584 52	XLHBOX8	Xenopus laevis XLHBOX	4.94e-01
11	28	2.3	2770 91	RNPFKL	Rat PFK-L MRNA for li	4.94e-01
12	28	2.3	38476 116	HUMU54B12	Human cosmid U54B12.	4.94e-01
13	28	2.3	38476 83	HUMU54B12	Human cosmid U54B12.	4.94e-01
14	28	2.3	41611 74	HSF77D12	Human DNA sequence fr	4.94e-01
15	28	2.3	117261 34	HS426N21	Human DNA sequence **	4.94e-01
16	28	2.3	242825 76	HSPEX	H.sapiens PEX gene.	4.94e-01
17	27	2.2	1108 46	BTINT6MR	B.taurus interleukin-	1.51e+00
18	27	2.2	1118 59	ATP22A	A.thaliana MRNA for p	1.51e+00
19	27	2.2	1160 91	RN17BHD1	R.norvegicus MRNA for	1.51e+00
20	27	2.2	1322 74	HSLUCAR	H.sapiens MRNA for no	1.51e+00
21	27	2.2	1717 90	MUSURSK	Mus musculus ultra hi	1.51e+00
22	27	2.2	1961 68	BTIMPP	S.tuberosum MRNA for	1.51e+00
23	27	2.2	3200 60	BNMYBIPRO	B.napus MRNA for myro	1.51e+00
24	27	2.2	36273 79	HUM7501	Homo sapiens chromoso	1.51e+00
25	27	2.2	44519 32	CED1086	*** SEQUENCING IN PRO	1.51e+00
26	27	2.2	105576 70	HS108120B	Human DNA sequence fr	1.51e+00
27	27	2.2	124065 34	HS440021	Human DNA sequence **	1.51e+00
28	27	2.2	135048 6	HSU91325	Human chromosome 16p1	1.51e+00
29	27	2.2	135048 116	HSU91325	Human chromosome 16p1	1.51e+00
30	27	2.2	215285 116	HSU91322	Human chromosome 16p1	1.51e+00
31	27	2.2	215285 6	HSU91322	Human chromosome 16p1	1.51e+00
32	26	2.1	354 8	OFU89259	Oxytricha fallax 57KD	4.51e+00
33	26	2.1	1341 91	RNRPA2	R.norvegicus MRNA for	4.51e+00
34	26	2.1	15000 117	HSRNGENE	H.sapiens ung gene fo	4.51e+00
35	26	2.1	18648 77	HSU18671	Human Stat2 gene, com	4.51e+00
36	26	2.1	28230 73	HSCG1160	Human cosmid CG1160 f	4.51e+00
37	26	2.1	54336 83	HUMXPDG1	Homo sapiens ERCC2 ge	4.51e+00
38	26	2.1	90638 34	HS6587	Human DNA sequence **	4.51e+00
39	26	2.1	129252 116	HSU91327	Human chromosome 16p1	4.51e+00
40	26	2.1	129252 6	HSU91327	Human chromosome 16p1	4.51e+00
41	26	2.1	167880 6	HSU91323	Human chromosome 16p1	4.51e+00
42	26	2.1	195158 34	HS473B4	Human DNA sequence **	4.51e+00
43	26	2.1	195167 122	HS431A14	Human DNA sequence **	4.51e+00
44	26	2.1	195167 109	HS431A14	Human DNA sequence **	4.51e+00
45	26	2.1	220633 6	HSU91321	Human chromosome 16p1	4.51e+00

ALIGNMENTS

RESULT 1 D84376 1212 bp mRNA ROD 03-OCT-1996
LOCUS House mouse; Musculus domesticus kidney mRNA for Phosphatidic acid
DEFINITION Phosphatase, complete cds.
ACCESSION D84376
NID g1487872
KEYWORDS Phosphatidic acid phosphatase; 35-kDa phosphatidic acid phosphatase.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1212)
AUTHORS Kanoh, H.
TITLE Direct Submission
JOURNAL Submitted (10-APR-1996) to the DDBJ/EMBL/GenBank databases. Hideo Kanoh, Sapporo Medical University School of Medicine, Department of Biochemistry, West-17, Sapporo, Hokkaido 060, Japan (E-mail: kanoh@serpent.cc.sapmed.ac.jp, Tel:011-611-2111(ex.2290), Fax:011-612-5861)


```

JOURNAL Development 104, 787-794 (1988)
REMARK revised by [2]
REFERENCE 2 (bases 1 to 1584)
AUTHORS Gama, L.
TITLE Direct Submission
JOURNAL Submitted (08-AUG-1994) Laura Gama, 615-343-8258 (lab) 615-343-4539
COMMENT (fax) e-mail: wrightr@ctr.vax.vanderbilt.edu
FEATURES Data kindly reviewed (22-JUN-1990) by De Roberts E.M.
source Location/Qualifiers
1..1584
/organism="Xenopus laevis"
/clone_lib="Charon 4A"
70..885
/notes="XlHbox 8 protein"
/codon_start=1
/db_xref="PID:g530783"
/db_xref="SWISS-PROT:PL4837"
/translation="MNADQYKPPQADYKPECAFORSQGDYNPSPACLYNGRQOQA
AYSNPLVADPGSPDPDISPYEPPPISEEPVPHLHHHHYHHHHHPGIPHPHHPFP
DNESTGLEERNTLLPFPWMSKTSKSHWGOWTDCSYIMEQENKRTTAYTRAQLL
ELEKEFLNKYISRPRVELAVMLNLTERRHIKIFONRRMKWKEEDKRGKSGDPEQ
DSVVSSADVLKDEPOCLGNSQTKDGLVLSPLPTSSQPNQVPSIGSLRQARR"
BASE COUNT 527 a 342 c 303 g 412 t
ORIGIN
Query Match 2.3%; Score 28; DB 52; Length 1584;
Best Local Similarity 79.2%; Pred. No. 4.94e-01;
Matches 38; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Db 1526 ttattattataaagcacatatgtgtaaaaaaataaaaaa 1573
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1185 TTTTGTAGTAATAGGCGACCTGTTTCACAAAAA 1232

RESULT 11
LOCUS RNPFLK 2770 bp RNA ROD 16-DEC-1991
DEFINITION Rat PFK-L mRNA for liver phosphofructokinase.
ACCESSION X58865
NID 956886
KEYWORDS phosphofructokinase.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
Murinae; Rattus.
REFERENCE 1 (bases 1 to 2770)
AUTHORS Hotta, K.
TITLE Direct Submission
JOURNAL Submitted (03-APR-1991) K. Hotta, Second Dept of Internal Medicine,
Osaka University Medical School, 1-1-50 Fukushima, Fukushima-ku,
Osaka 553, JAPAN
REFERENCE 2 (bases 1 to 2770)
AUTHORS Hotta, K., Nakajima, H., Yamasaki, T., Hamaguchi, T., Kuwajima, M.,
Noguchi, T., Tanaka, T., Kono, N. and Tarui, S.
TITLE Rat-liver-type phosphofructokinase mRNA. Structure, tissue
distribution and regulation
JOURNAL Eur. J. Biochem. 202 (2), 293-298 (1991)
MEDLINE 92104147
FEATURES Location/Qualifiers
source 1..2770
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/tissue_type="liver"
/clone_lib="lambda-gt10"
/clone="lambda-PFKL1-PFKL8"
1..2740
/evidence="experimental"
/gene="PFK-L"
62..70
/gene="PFK-L"
67..2409
/gene="PFK-L"
/EC_number="2.7.1.11"

```

```

/codon_start=1
/product="6-phosphofructokinase"
/db_xref="PID:g56887"
/db_xref="SWISS-PROT:P30835"
/translation="MATVDLEKLRMSGAGKAGLVITSGDAGQGNAAVAVRVMGIYV
GAKFLITYEGYGLVEGGENIKPANLWSVNIQLGGTIIGSARC KAFTRREGRLAA
YNLQHGTTNLCVIGGDSGLTGANFRNWSLLELVKEGKISESTAQNTAHLISAG
LVGSDNDFCGTDMTIGTDSALHRIMEVIDAITTAQSHQRTFVLEVMGRHCGYLALV
SALASGADMLFPEAPPEDGHNEMCERLGETRSRGSRLNIIIAEGRIDRHKPISS
SYKDLVQRLGFDTRVTVLHVQRGTPSFADRVLSKMGNEAVMALLETPTDPAC
VYLSGNSVRLPLMECVTKDQVOKANDERDEAIQLRGSRPENNKKIYKLLAHQK
VSKEKSNFSLAINVGAAGNAAVSAVETGISEGHTVTVVHDGFEAGKQVQEV
GWHDVAGWLGSGGMLGTRKTLPKPLHEAIVENLRTYNIHALLVIGGFEASGTRRV
EARGRYEELCIVMCVIPATISNNVPGTDFSLGSDTAVNAAMESCDRIKQASGTRRV
FIVETMGYCGYLATVTCIAGADAAYVDFPNHDLKANVEHTEKMKTDIQRLV
LANEKCHEYTFEFLNLYSSEGRGVFDCRNVLHQQGGAPTFDRNYGTKGVKA
MLMSEKLRDVRKRGVFANAPDSACVIGLRKKVVAFSVTELEKTKETDFEHRMPREQ
WLNLRMLMLAHYRISMADYVSGLEHVTRRTLSIDKGF"
2722..2727
/gene="PFK-L"
2740
/polya_signal
polya_site
BASE COUNT 609 a 762 c 864 g 535 t
ORIGIN
Query Match 2.3%; Score 28; DB 91; Length 2770;
Best Local Similarity 79.2%; Pred. No. 4.94e-01;
Matches 38; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Db 2713 tttttagaataaagcacctgttttagaaaaaataaaaaa 2760
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1185 TTTTGTAGTAATAGGCGACCTGTTTCACAAAAA 1232

RESULT 12
LOCUS HUM054B12 38476 bp DNA PRI 12-FEB-1997
DEFINITION Human cosmid U54B12, complete sequence.
ACCESSION U73024
NID g1613887
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 38476)
AUTHORS Fulton, B. and Rohlfing, T.
TITLE The sequence of H. sapiens cosmid U54B12
JOURNAL Unpublished (1996)
REFERENCE 2 (bases 1 to 38476)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (01-OCT-1996)
REFERENCE 3 (bases 1 to 38476)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (15-OCT-1996)
COMMENT Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
e-mail: sapiens@watson.wustl.edu

```

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was

(TM)

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	83	6.7	1047	2	Q10572	Human Natriuretic Pep	1.19e-25
c	75	6.1	1047	2	Q10572	Human Natriuretic Pep	2.69e-25
3	70	5.7	74	20	T23564	Human gene signature	1.32e-22
c	42	3.4	91	9	Q51746	Oligonucleotide probe	3.66e-08
5	42	3.4	91	9	Q51746	Oligonucleotide probe	3.66e-08
c	6	4.2	204	1	N81164	Base substituted E.co	3.66e-08
7	40	3.2	204	1	N81164	Base substituted E.co	3.41e-07
8	36	2.9	114	12	Q70457	Generic DNA sequence	2.71e-05
9	34	2.8	114	12	Q70465	Generic DNA sequence	2.29e-04
10	34	2.8	114	12	Q70466	Generic DNA sequence	2.29e-04
11	34	2.8	114	12	Q70469	Generic DNA sequence	2.29e-04
12	34	2.8	114	12	Q70468	Generic DNA sequence	2.29e-04
c	13	34	114	12	Q70468	Generic DNA sequence	2.29e-04
c	14	34	114	12	Q70469	Generic DNA sequence	2.29e-04
15	33	2.7	81	21	T13611	DC43 TSAR library	6.57e-04

```
PF 22-JUN-1990; U03586.
PR 23-JUN-1989; US-370673.
PA (GETH ) GENENTECH INC.
PI Chang M, Goeddel D, Lowe D;
DR WPI: 91-036711/05.
DR N-PSDB; Q10324.
PT Natriuretic protein receptor B - for diagnosis and treatment of
PS Claim 3; Fig 1; 49pp; English.
CC The sequence was derived from the DNA encoding natriuretic peptide
CC receptor B, NPRB, having guanyl cyclase (GC) activity and protein
CC kinase activity. The DNA can be inserted into expression vectors
CC for the prodn. of the protein, opt. after being mutated to produce
CC NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr=
CC 114,952). The protein (or variants) can be used in treatment of
CC natriuretic peptide disorders, and also to isolate peptides using
CC affinity chromatography. Antibodies with affinity for NPRB can
CC also be prep'd.
SQ Sequence 1047 BP; 87 A; 15 C; 83 G; 51 T;

Query Match 6.7%; Score 83; DB 2; Length 1047;
Best Local Similarity 8.5%; Pred. No. 1.19e-29;
Matches 72; Conservative 236; Mismatches 530; Indels 7; Gaps 7;

Db 194 snsvnhvnyrnnggnnathnrrngrvncgnnnmhnmmhnnnnnntngdyvnn 253
QY 27 GCTGCTCGAGGTGCTGCTACTGTGCTGCTCCTCCCTGCTGCTATCTGCTGAGCTGGT 86
Db 254 ydvngnsnragtrntrgrndrtrnnananrnanntvntvtyrnnnnnnnnnnrn 313
QY 87 GAACGCCCCGTACAGCGAGGATTTTACTCGGGGATGACTCCATCCGGTACCCCTACC 146
Db 314 nraerdngvngnsnmnnagcnydgnnyanvnnntnnngtrndgnrnvnmkgrry 373
QY 147 TCCAGATACCATCACCCACGGGCTCATGGCTGGGTACCATCAGGCCACCGTCATCCT 206
Db 374 hgvtnvnmkndrntdvnwmgdndsgdnnnaahysganknnwvtrnnnnwkgann 433
QY 207 TGTCTCGCGCGGGAA-GCCTACTGCTGTACACAGACCGGCTGTATCTCGCTCGGACT 265
Db 434 -sdnnncandndsdsktstnanvangtntnmmgvssnnnnrknmmknasmnw 492
QY 266 TCAACAACACTAGTGGCTGCTGTATACAAAGTGTCTGGGACCTTCTGTTGGGGCTGCC 325
Db 493 rnrwnnnngnsnryhkgagrsntnrgssygsnmtahckynnnantghnknvvanhk 552
QY 326 TGACCCAGTCTCTGACAGACTGCGCCAGGTACATGAT-TGGGGGT-CTGAAGCCCACTT 383
Db 553 vnkrnntnrnvnnnnkhmrdivnnhnrtnngacndnann-cnvtnycnrgsnnndnnd 611
QY 384 CCTAGCCGTCTGCGACCCCGACTGGAGCCGGGTCAACTGCTCGTCTATGTGAGCTGGA 443
Db 612 snndwmrvsnndnvkgnannhnsnshgsknsncvdsrvnknktdygnasrst 671
QY 444 GAAGGTGTGACGGGAAACCCCTGCTGATGTACCCAGGCGGAGGTGTCTTTCTACTCGGG 503
Db 672 anndnnanvknntnannngsnnttgmnaadvysngnnnnnnnnnnrsgnnyngndn 731
QY 504 ACACCTCTCTTGGGATGTACTGCATGGTGTCTTGGCGCTGTATGTGAGGACGACT 563
Db 732 snknvkvngnrnyrnarnrdtrnnnnnnnnnnmrcwandnarnndngnknknrrnnk 791
QY 564 CTGTTGAAGTGGSCAGGCTGTGCGACCCACAGTCCAGTCTTCTCTGTTGGGCT-TTG 622
Db 792 nggtssndnnnnmnyannnnknvnrtrtaaynnknkkanannynnnnnhavannkrgnt 851
QY 623 CCCTCTACGTGGGTACACCGCGTGTCTGATTACAAACACCACTGAGGAGGTCTCTTG 682
Db 852 vnanandsvtynsdvngntansnstnmvvtvtnndnntcndann-dndvvykvntngd 910
QY 583 TTGGCCCTCTGACGGGGGACTGTGGCTGCCCTCCTACTGTCTGTACATCTCAGACTCT 742
Db 911 aymvsvngnrgnrhannhannmanannndavssnrnrhrhndnnrnrgvhtgncagvv 970

QY 743 TCAAAGCCGACCCCAACAGCAGCTGTCTGAAGGAGGAGCTGGAACGAGCCAGCC 802
Db 971 gknmrycngngdvtntasrmnngnanknhvsttkdandnngcnnnnrgrdvnmkgk 1030
QY 803 TGTCACTGACGTTGACCTGGGGGAGGCTGACCAACCACTTATGGGATACCCGCACT 862
Db 1031 mrtv 1035
QY 863 CTCTCT 867

RESULT 2
ID Q10572 standard; DNA; 1047 BP.
AC Q10572;
DT 09-APR-1991 (first entry)
DE Human Natriuretic Peptide Receptor B.
KW NPRB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
KW hyperaldosteronism; glaucoma; guanyl cyclase.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= signal sequence
FT Protein 12
FT /label= mature NPRB
FT Domain 23..455
FT /label= extracellular domain
FT /note= "binds natriuretic peptides A,B and C]"
FT Domain 456..456
FT /label= transmembrane domain
FT Domain 479..1047
FT /label= cytoplasmic domain
FT /note= "GC and protien kinase activity"
FT Modified-site 24..26
FT /label= N-glycos_site
FT Modified-site 35..37
FT /label= N-glycos_site
FT Modified-site 161..163
FT /label= N-glycos_site
FT Modified-site 195..197
FT /label= N-glycos_site
FT Modified-site 244..246
FT /label= N-glycos_site
FT Modified-site 277..279
FT /label= N-glycos_site
FT Modified-site 349..351
FT /label= N-glycos_site
FT Modified-site 600..602
FT /label= N-glycos_site
PN W09100292-A.
PD 10-JAN-1991.
PF 22-JUN-1990; U03586.
PR 23-JUN-1989; US-370673.
PA (GETH ) GENENTECH INC.
PI Chang M, Goeddel D, Lowe D;
DR WPI: 91-036711/05.
DR N-PSDB; Q10324.
PT Natriuretic protein receptor B - for diagnosis and treatment of
PS Claim 3; Fig 1; 49pp; English.
CC The sequence was derived from the DNA encoding natriuretic peptide
CC receptor B, NPRB, having guanyl cyclase (GC) activity and protein
CC kinase activity. The DNA can be inserted into expression vectors
CC for the prodn. of the protein, opt. after being mutated to produce
CC NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr=
CC 114,952). The protein (or variants) can be used in treatment of
CC natriuretic peptide disorders, and also to isolate peptides using
CC affinity chromatography. Antibodies with affinity for NPRB can
CC also be prep'd.
SQ Sequence 1047 BP; 87 A; 15 C; 83 G; 51 T;

Query Match 6.1%; Score 75; DB 2; Length 1047;
Best Local Similarity 8.6%; Pred. No. 2.69e-25;
```


[illegible][illegible]

PT Identifying proteins or peptide(s) which bind a ligand - by
 PT screening a recombinant vector library expressing fusion proteins
 PT comprising a binding domain and an effector domain
 PS Disclosure, Page 35; 255pp; English.
 CC Q70469 is a generic DNA sequence used to generate random TSAR peptide
 CC This generic formula can be represented as follows: X(TGC)(NNB)10-
 CC (TGC)(NNB)62(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking restriction
 CC sites (X is not the same as Y) that are not specified further. This
 CC sequence generates peptides that are cloverleaf in structure. Other
 CC generic sequences are shown in Q70465-68. Other specific peptides
 CC generated by these generic sequences are shown in R65150-54. TSARs are
 CC concatenated heterofunctional proteins or peptides, comprising at least
 CC two functional regions - a binding domain with affinity for a ligand and
 CC a second effector peptide portion that is chemically or biologically
 CC active. They may further comprise a linker peptide between the 2 domains.
 CC The oligonucleotides are also designed so that the expressed peptide
 CC contains 2 or 4 cysteine residues positioned in, or flanking, the
 CC unpredicted or variant residues, these residues confer some degree of
 CC conformational rigidity to the peptides. The TSARs or compps. comprising

M P S R C H
***** (TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPSrch_n n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Thu Nov 6 12:46:31 1997; MasPar time 638.42 Seconds
Tabular output not generated.

Title: >US-08-842-827-7
Description: (1-1232) from US08842827.seq
Perfect Score: 1232
N.A. Sequence: 1 ACCATGCGAGCGAGTGGGT.....CAAAAAAAAAAAAAAAAA 1232
Comp: TGGTACGTCGCCACCA.....GTGTTTTTTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 707517 seqs, 256659390 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: EST-STS
1:EST1 2:EST2 3:EST3 4:EST4 5:EST5 6:EST6 7:EST7 8:EST8
9:EST9 10:EST10 11:EST11 12:EST12 13:EST13 14:EST14
15:EST15 16:EST16 17:EST17 18:EST18 19:EST19 20:EST20
21:EST21 22:EST22 23:EST23 24:EST24 25:EST25 26:EST26
27:EST27 28:EST28 29:EST29 30:EST30 31:EST31 32:EST32
33:EST33 34:EST34 35:EST35 36:EST36 37:EST37 38:EST38
39:EST39 40:EST40 41:EST41 42:EST42 43:EST43 44:EST44
45:EST45 46:EST46 47:EST47 48:EST48 49:EST49 50:EST50
51:EST51 52:EST52 53:EST53 54:EST54 55:EST55 56:EST56
57:EST57 58:EST58 59:EST59 60:EST60 61:EST61 62:EST62
63:EST63 64:EST64 65:EST65 66:EST66 67:EST67 68:EST68
69:EST69 70:EST70 71:EST71 72:EST72 73:EST73 74:EST74
75:EST75 76:EST76 77:EST77 78:EST78 79:EST79 80:EST80
81:EST81 82:EST82 83:EST83 84:EST84 85:EST85 86:EST86
87:EST87 88:EST88 89:EST89 90:EST90 91:EST91 92:EST92
93:EST93 94:EST94 95:EST95 96:EST96 97:EST97 98:EST98
99:EST99

Database: EST-STS-TWO
100:EST100 101:EST101 102:EST102 103:EST103 104:EST104
105:EST105 106:EST106 107:EST107 108:EST108 109:EST109
110:EST110 111:EST111 112:EST112 113:EST113 114:EST114
115:EST115 116:EST116 117:EST117 118:EST118 119:EST119
120:EST120 121:EST121 122:EST122 123:EST123 124:EST124
125:EST125 126:EST126 127:EST127 128:EST128 129:EST129
130:EST130 131:EST131 132:EST132 133:EST133 134:EST134
135:EST135 136:EST136 137:EST137 138:EST138 139:EST139
140:EST140 141:EST141 142:EST142 143:EST143 144:EST144
145:EST145 146:EST146 147:EST147 148:EST148 149:EST149
150:EST150 151:EST151 152:EST152 153:EST153 154:EST154
155:EST155 156:EST156 157:EST157 158:EST158 159:EST159
160:EST160 161:EST161 162:EST162 163:EST163 164:EST164
165:EST165 166:EST166 167:EST167 168:EST168 169:EST169
170:EST170 171:EST171 172:EST172 173:EST173 174:EST174

175:EST175 176:EST176 177:EST177 178:EST178 179:EST179
180:EST180 181:EST181 182:EST182 183:EST183 184:EST184
185:EST185 186:EST186 187:EST187 188:EST188 189:EST189
190:EST190 191:EST191 192:EST192 193:EST193 194:EST194
195:EST195 196:EST196 197:EST197 198:EST198

Statistics: Mean 11.465; Variance 3.184; scale 3.601

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description	Pred. No.
1	520	42.2	610	189	AA151344	z125a07.r1 Soares pre	0.00e+00
2	396	32.1	495	189	AA149244	z125a07.s1 Soares pre	0.00e+00
3	375	30.4	404	160	AA042850	zk52d08.r1 Soares pre	0.00e+00
4	339	27.5	349	105	HSC1GE101	H. sapiens partial CD	0.00e+00
5	297	24.1	413	21	R34980	yh86h03.r1 Homo sapie	0.00e+00
6	277	22.5	367	178	AA074521	zml7h03.s1 Stratagene	0.00e+00
7	255	20.7	264	105	HSC1GE102	H. sapiens partial CD	0.00e+00
8	195	15.8	237	21	R34878	yh86h03.s1 Homo sapie	3.44e-33
9	161	13.1	167	137	AA044400	zk52d08.s1 Soares pre	2.98e-183
10	162	13.1	177	162	AA102178	zml7h03.r1 Stratagene	7.99e-185
11	105	8.5	511	154	AA008988	mg99e04.r1 Soares mou	2.94e-103
12	97	7.9	423	78	R97295	yq74h06.r1 Homo sapie	3.95e-92
13	95	7.7	490	92	R31047	yx51h07.r1 Homo sapie	2.29e-89
14	75	6.1	446	29	R63796	y115h04.r1 Homo sapie	3.28e-62
15	72	5.8	472	114	W04968	za43g09.r1 Soares fet	3.14e-58
16	65	5.3	357	159	AA037575	zk43f08.r1 Soares pre	4.59e-49
17	61	5.0	346	122	W61841	md35a11.r1 Life Tech	6.48e-44
18	60	4.9	263	145	R75377	MDB0538R Mouse brain,	1.22e-42
19	60	4.9	480	169	W45113	zk21c09.r1 Soares sen	1.22e-42
20	59	4.8	170	189	AA152122	z141g01.r1 Soares pre	2.29e-41
21	59	4.8	331	158	AA033777	zk19g11.r1 Soares pre	2.29e-41
22	59	4.8	398	138	AA048371	mj28d11.r1 Soares mou	2.29e-41
23	59	4.8	446	171	W67666	zk38a11.r1 Soares fet	2.29e-41
24	59	4.8	456	169	W45106	zk21a10.r1 Soares sen	2.29e-41
25	59	4.8	456	67	H68363	yr82f11.r1 Homo sapie	2.29e-41
26	59	4.8	471	189	AA152123	z141g02.r1 Soares pre	2.29e-41
27	59	4.8	474	113	W01275	za40f11.r1 Soares fet	2.29e-41
28	58	4.7	356	165	C17623	Human placenta cDNA 5	4.22e-40
29	57	4.6	348	54	H04659	yf49a08.r1 Homo sapie	7.71e-39
30	57	4.6	351	85	H57213	yr08d07.r1 Homo sapie	7.71e-39
31	57	4.6	385	88	H66929	yr86a11.r1 Homo sapie	7.71e-39
32	55	4.5	208	165	C16881	Human placenta cDNA 5	2.47e-36
33	55	4.5	449	173	W70040	zk49b06.s1 Soares fet	2.47e-36
34	55	4.5	451	83	H50533	yp08a11.s1 Homo sapie	2.47e-36
35	55	4.5	452	85	H54373	yq93g09.r1 Homo sapie	2.47e-36
36	55	4.5	555	161	AA056693	zk81b05.s1 Soares pre	2.47e-36
37	56	4.5	572	172	AA058383	zk81b05.r1 Soares pre	1.39e-37
38	54	4.4	545	137	AA043085	zk48b02.r1 Soares pre	4.35e-35
39	51	4.1	370	30	R71019	y150a11.r1 Homo sapie	2.17e-31
40	50	4.1	479	74	H90961	yus6e01.r1 Homo sapie	3.60e-30
41	50	4.1	485	60	H12562	yf12d10.r1 Homo sapie	3.60e-30
42	49	4.0	292	20	R00690	ye74a10.r1 Homo sapie	5.87e-29
43	47	3.8	422	4	T69858	yc18c09.r1 Homo sapie	1.49e-26
44	45	3.7	287	142	N86994	Li1773F Fetal heart, L	3.51e-24
45	45	3.7	546	171	W67667	zk38a11.s1 Soares fet	3.51e-24

ALIGNMENTS

1 AA151344 610 bp mRNA EST 10-DEC-1996
LOCUS z125a07.r1 Soares pregnant uterus NDHPU Homo sapiens cDNA clone
DEFINITION 502932 5',
ACCESSION AA151344
NID g1719889
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 610)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 452.

FEATURES
source
Location/Qualifiers
1..610
/organism="Homo sapiens"
/note="Organ: uterus; Vector: pT7T3-Pac; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTGGAAGAATCGCGCGCCGCTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."
/clone="502932"
/clone_lib="Soares pregnant uterus NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
<1..5610

BASE COUNT 132 a 172 c 171 g 132 t 3 others
ORIGIN

Query Match 42.2%; Score 520; DB 189; Length 610;
Best Local Similarity 96.4%; Pred. No. 0.00e+00;
Matches 590; Conservative 0; Mismatches 12; Indels 10; Gaps 10;

Db 2 ttgcccctctacgtggg-tacacccggtgtctgattacaaacaccactggagcgatgtcc 60
QY 620 TTGCGCTCTACGTGGGTACACCGCGGTGCTGATTACAAACACCAGCTGGAGCGATGCC 679
Db 61 ttgtggcctctcagggggnactgggtgctccctcactgtctgtaacatctcagact 120
QY 680 TTGTGGCTCTCTGTCAGGGGGCACTGGTGGCTGCCCTCACTGTGCTACATCTCAGACT 739
Db 121 tcttcaagcccgaccccccacagactgtctgaagagagagctggaacggaagccca 180
QY 740 TCTTCAAGCCCGACCCCGACACAGCTGTCTGAAGAGGAGGAGCTGGAAACGGAAGCCCA 799
Db 181 gctgtcactgactgtgacccctggg-cgaggtctgacacacacacact-atgg-atagccgcg 237
QY 800 GCCTGTCACTGAGCTTGACCCCTGGCGAGGCTGACCACACACACATATGGGATACCCGC 859
Db 238 actcctctctcagcccgagcccccaccccaagcagggagctactgtgagtcagctga 297
QY 860 ACTCTTCTTCTGAGCGCGGACCCGCCCA-GGCA-GGGAGCTGCTGTGAGTCCAGCTGA 917
Db 298 ggcaccccaagtggtccctccaaagccctgttaggcactgagggctctggagcggtc 357
QY 918 TGCCCCACCCA-GGTGGTCCCTCCA-GCC-TGGTTAGCACTGAGGGTCTCTGACGGGCTC 974

Db 358 caggaaacctggctgatgggagcagtgagcgggtccctgctcccccctgcctgactgg 417
QY 975 CAGGAACCCCTGGCTGATGGGAGCAGTAGCGGT-TCCGCTGCCCTGCCTGCCTGCTG 1033
Db 418 acacagattctggagatgctgggtagcctcagcatttgggggggaacctgttcccgctc 477
QY 1034 ACCAGAGGCTCGAGATGCTGGTAGCCCTCAGCATTTGGAGGGGAACCTGTTCCCGTC 1093
Db 478 ggtccccaataatcncctctcttttgggttaagggaagggcagagacagatcagatag 537
QY 1094 GGTCCCAAAATATCCCTTCTTTTATGGGGTTAAGGAAGGAGCCAGAGATCAGATAG 1153
Db 538 tctgtttttaaataatgtaagtataggtgttttttagtaaaatagggcacctgttttac 597
QY 1154 TCGTGTTTTGTAAATGTAATCTATATGTGTTTATTAGTAAATAGGGCACCTGTTTCAC 1213
Db 598 aaaaaaataaaaaa 609
QY 1214 AAAAAAAAAAAAAA 1225

RESULT 2
LOCUS AA149244 495 bp mRNA EST 10-DEC-1996
DEFINITION z125a07.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 502932 3'
ACCESSION AA149244
NID Q1719838
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 495)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 336.
Location/Qualifiers
1..495
/organism="Homo sapiens"
/note="Organ: uterus; Vector: pT7T3-Pac; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTGGAAGAATCGCGCGCCGCTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."
/clone="502932"
/clone_lib="Soares pregnant uterus NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
complement(<1..>495)

BASE COUNT 106 a 135 c 140 g 109 t 5 others
ORIGIN

FEATURES
source

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 367)

TITLE Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

JOURNAL WashU-Merck EST Project

COMMENT Unpublished (1995)

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 194.

FEATURES
source Location/Qualifiers
1..367
/organism="Homo sapiens"
/note="Organ: pancreas; Vector: pBluescript SK-; Site: 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Pancreatic adenocarcinoma cell line. Average insert size: 1.0 Kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGGAG 3' -3' adaptor sequence: 5' CTCGAGTGTGTTTTTTTTTTT 3'"
/clone="525941"
/clone_lib="Stratagene pancreas (#937208)"
/lab_host="SOLR cells (Kanamycin resistant)"
complement(1..>367)
89 a 107 c 97 g 70 t 4 others

BASE COUNT
ORIGIN

Query Match 22.5%; Score 277; DB 178; Length 367;
Best Local Similarity 96.2%; Pred. No. 0.00e+00;
Matches 328; Conservative 0; Mismatches 5; Indels 8; Gaps 7;

Db 1 tgtgaacagggtgcctattttactataaaacacacatacatattttacaaacagc 60
|||||
Cp 1214 TGTGAACAGGTGCCCTATTACTATAAACACACATATACATATTACAAACAGC 1155
|||||

Db 61 aactatctgatctctggttccttccttaacccataaaagaagggtatttggggac 120
|||||
Cp 1154 AACTATCTGATCTCGGTCCCTTCCTTAACCCATAAAAGAGGGGATATTGGGGAC 1095
|||||

Db 121 cgacgggaacaggttccctccaaatgctgagggctaccacaggtatctccagactctg 180
|||||
Cp 1094 CGACGGGAACAGGTTCCTCCAAATGCTGAGGGCTACCCAGGCATCTCCAGACTCTGG 1035
|||||

Db 181 tcacgtgcaggcagggcagcagagctgctcactgtccatccatcagccagggttct 240
|||||
Cp 1034 TCAGTGTGAGGCGAGGGGAGCGGAACC-GTCACTGTCTCCATCAGCCCCAGGGTTCCT 976
|||||

Db 241 ggagccggtccagaagccctcagtcagtcctcaancaaagggtggaaggaccactgngtgg 300
|||||
Cp 975 GGAGCCGTCACAGAA-CCCTCAGTGCCTAACCA--GGCTTGA-GGGACCACCTGGTGGG 920
|||||

Db 301 cctcaagctgggaactcaagcagctcccttgccttgcctggggccgg 341
|||||
Cp 919 CATCA-GCTGT-AGTCACAGCAGCTCCCT-GCCTGGGGCGGG 882
|||||

RESULT 7
LOCUS HSC1GE102 264 bp RNA EST 21-SEP-1995
DEFINITION H. sapiens partial cDNA sequence; clone c-1ge10.
ACCESSION Z39680

9565475
partial cDNA sequence; transcribed sequence fragment.
human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 264)

TITLE Direct Submission

JOURNAL Submitted (24-OCT-1994) Genethon, B.P. 60, 91002 Evry Cedex France and Genetique Moleculaire et Biologie du developpement, CNRS UPR420 B.P. 8, 94801 Villejuif Cedex France.E-mail: genexp@genethon.fr

COMMENT 2 (bases 1 to 264)
Genexpres.
The Genexpres cDNA program
Unpublished
3 (bases 1 to 264)
Aufray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Devignes, M.D., Duprat, S., Houlgatte, R., Jumeau, M.N., Lamy, B., Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pietu, G., Pouliot, X., Sebastiani-Kabaktchis, C. and Tessier, A.
IMAGE: molecular integration of the analysis of the human genome and its expression
C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
95277534
Clone library from B.Souares, Psychiatry Dept. Columbia University USA:

Cloning_method: total mRNA was oligo-(dT) primed and directionally cloned 5' -> 3' into the HindIII -> NotI sites of the laimID BA vector;
Sequencing_method: single read, full automatic;
Primer: (-21)M13 universal;
cDNA sequence complementary to mRNA (3'end)
Stretch_removed: 18 T removed at sequence 5'end
Normalization_method: Bento Soares, P.N.A.S in press;
Genexpres_library_id: C;
Genexpres_sequence_id: alc-lge10;

No significant homology found with :
genbank release 81 swissprot release 28.
Location/Qualifiers
1..264
/organism="Homo sapiens"
/isolate="muscular atrophy patient"
/dev_stage="3 months old"
/tissue_type="total brain"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
66 a 79 c 61 g 54 t 4 others

BASE COUNT
ORIGIN

Query Match 20.7%; Score 255; DB 105; Length 264;
Best Local Similarity 98.1%; Pred. No. 0.00e+00;
Matches 256; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 4 aanaaggtccctatttactataaaacacatacatattttacaaacagcaacta 63
|||||
Cp 1209 AACAGGTGCCCTATTATTACTAAAAACACATATACATTACATTTTACAAACAGCAACTA 1150
|||||

Db 64 tctgatctctggtcccttccttaacccataaaagaagggtatttggggaccgacg 123
|||||
Cp 1149 TCTGATCTCTGGTCCCTTCCTTAACCCCATAAAAAGAGGGGATATTGGGGACCGAGC 1090
|||||

Db 124 ggaacaggttccctccaaatgctgagggctaccacagcatctccagactctctggtccag 183
|||||
Cp 1089 GGAACAGGTTCCCTCCAAATGCTGAGGGCTACCCAGGCATCTCCAGACTCTCTGGTCCAG 1030
|||||

Db 184 tgcagggcagggggcagcgggancctcactgtcccatcagccagggttctctggagcc 243
|||||
Cp 1029 TCGAGGGAGGGGGGAGCGGAACCGCTCACTGCTCCCATCAGCCAGGGTTCCTGGAGCC 970
|||||

Db 244 cgtccagagccctcagtgct 264

[illegible]


```

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
AUTHORS Vertebrata; Euthera; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 177)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevisan,E.,
Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
WashU-Merck EST Project
WashU-Merck EST Project
Unpublished (1995)

TITLE
JOURNAL
COMMENT Contact: Wilton RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 100.
Location/Qualifiers
FEATURES
source
1..177
/organism="Homo sapiens"
/Note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Pancreatic adenocarcinoma cell line. Average
insert size: 1.0 Kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGCGACGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTCTTTTCTTTT 3'"
/clone="525941"
/clone.lib="Stratagene pancreas (#37208)"
/lab.host="SOLR cells (kanamycin resistant)"
<1..>177
BASE COUNT 35 a 56 c 44 g 38 t 4 others
ORIGIN
Query Match 13.1%; Score 162; DB 162; Length 177;
Best Local Similarity 96.0%; Pred. No. 7.99e-185;
Matches 170; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
Db 1 aggatttactcggtggtgactccatccgtaccaccatccgtccagatcaccaccc 60
|||||
QY 105 AGGATTTTACTCGGGGAGTACTCCATCGGTACCC-TACCGTCAGATACCATCACCC 163
|||||
Db 61 acgggctcgtcggtggtggtcaccatcaccggtcaccgtccttctcgtcggtgggaan 120
|||||
QY 164 ACGGGCTCATGCTGGGGTCCACCATCACGGCCACCGTCATCTCTCGGCGGGGAAG 223
|||||
Db 121 cctactggtgtacacagacgggtctantctcgtcggtggtggtggtggtggtggt 177
|||||
QY 224 CCTACTGCTGTACACAGACCGGCTCTATTCTCGCTCGGACTTCAACAACACTACG 280
|||||

RESULT 11
LOCUS AA008988 511 bp mRNA EST 26-JUL-1996
DEFINITION mg99e04.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
clone 441150 5' similar to WP:T28D9.3 C502068 ;.
ACCESSION AA008988
NID g1465044
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euthera; Rodentia; Sciurognathi; Myomorpha; Muridae;
Murinae; Mus.
REFERENCE 1 (bases 1 to 511)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:266486
Possible reversed clone: similarity on wrong strand
Seq primer: ETPPrimer
High quality sequence stop: 325.
Location/Qualifiers
FEATURES
source
1..511
/organism="Mus musculus"
/strain="C57BL/6J"
/Note="Vector: pMT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCGAAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru KO, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pMT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M. Patricia Bonaldo."
/clone="441150"
/clone.lib="Soares mouse embryo NBME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab.host="DH10B"
<1..>511
BASE COUNT 129 a 127 c 123 g 132 t
ORIGIN
Query Match 8.5%; Score 105; DB 154; Length 511;
Best Local Similarity 69.1%; Pred. No. 2.94e-103;
Matches 190; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
Db 207 ttgaagaatcggtacatacaaaagcaaccagtagtcacatagcagctccctgaatgagt 266
|||||
Cp 746 TTGAAGAAGTCTGAGATGTAGCAGACAGTGAGGGGAGCCACCAGTGCCCTGCAGGAGG 687
|||||
Db 267 ccaactgtgagctcactccagtggtgtgttagtcagacactcagagaaagggccacatat 326
|||||
Cp 686 CCAACAAGGACATCGCTCCAGTGTGTGTGTAATCAGACACGCGGGTGTAGCCACGATAG 627
|||||
Db 327 atggaaaaagaagaatgagcccaaatgagcagtggtgtcgaagagcttgcagctctccc 386
|||||
Cp 626 AGGGCAAGAGCCACCAGGAAGAACCTGGACTGTGGTGTGCGAGCAGCGCGCTCCACTCC 567
|||||
Db 387 ttcatcctggcttgaagataaaagtgcgcaaacacagcatcagtagatagagaatgaagag 446
|||||
Cp 566 CAGAGTCGTGCTGCACATACAGCGCCCAAGAACACCATCATCATATCCCAAGGAAGAG 507
|||||
Db 447 tgtcccgagtagaagaacaaacctgcctccttgac 481
|||||
Cp 506 TGTCCTCGAGTAAAGACAAACCTGCCTCGGTGAC 472
|||||

RESULT 12
LOCUS R97295 423 bp mRNA EST 11-SEP-1995
DEFINITION YG74h06.r1 Homo sapiens cDNA clone 201563 5'.
ACCESSION R97295

```

```

NID          g982955
KEYWORDS     EST.
SOURCE       human clone=201563 primer=M13RP1 library=Soares fetal liver spleen
             1NFLS vector=pr773D (Pharmacia) with a modified polylinker
             host-DH10B (ampicillin resistant) Rsite1-Pac I Rsite2-Eco RI Liver
             and spleen from a 20 week-post conception male fetus. 1st strand
             cDNA was primed with a Pac I - oligo(dT) primer [5',
             AACTGGAGAATTAATTAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded
             cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
             I and cloned into the Pac I and Eco RI sites of the modified pr773
             vector. Library went through one round of normalization. Library
             constructed by Bento Soares and M.Fatima Bonaldo.
ORGANISM      Homo sapiens
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Homidae; Homo.
REFERENCE     1 (bases 1 to 423)
AUTHORS       Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
               Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
               Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
               Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
               Wilson,R.
TITLE         The WashU-Merck EST Project
JOURNAL       Unpublished (1995)
COMMENT       Contact: Wilson RK
               WashU-Merck EST Project
               Washington University School of Medicine
               4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
               Tel: 314 286 1800
               Fax: 314 286 1810
               Email: est@watson.wustl.edu
               High quality sequence stops: 301
               Source: IMAGE Consortium, LLNL
               This clone is available royalty-free through LLNL; contact the
               IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES      source
               1..423
               /organism="Homo sapiens"
               /clone="201563"
               <1..>423
BASE COUNT    108 a 88 c 103 g 117 t 7 others
ORIGIN
Query Match 7.9%; Score 97; DB 78; Length 423;
Best Local Similarity 67.2%; Pred. No. 3.95e-92;
Matches 182; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
Db 12 gaagggcaggtgtgtcttctattcaggncaactcttcgtttncatgtactgcagctgttt 71
QY 478 GAGGCCAGGTGTCTTCTACTCGGACACTCTTCTTGGGATGTACTGCATGGTCTC 537
Db 72 gtggcactttatcttcaagccagatgaagggagactgggcaagactcttaccnccaca 131
QY 538 TTGGCGTGTATGTATGACAGGACAGACTCTGTGTGAAGTGGGACGCTGCTGCGACCCACA 597
Db 132 ctgcaatttggtctgttgcgtatcattatgtatggcctttctcagttctcagttat 191
QY 598 GTCCAGTCTTCTCTGGTGGCCCTTGGCCCTCTACGTGGGCTACACCCGGGTCTGTATAC 657
Db 192 aaacaccactggagcagatgtgttgactggactcattcaggagagctcgtgttgcaatatta 251
QY 658 AAACACCACCTGGAGCGATGTCTTGTGGCTCTCTGCGAGGGGACACTGGTGGCTGCCCTC 717
Db 252 gttgctgtatgtatcggattcttccaaag 282
QY 718 ACTGTCTGCTACATCTCAGACTTCTTCAAAG 748

RESULT 13 N31047 490 bp mRNA EST 10-JAN-1996
LOCUS yx51h07.r1 Homo sapiens cDNA clone 265309 5' similar to WP:T28D9.3
DEFINITION

```

```

CE02068 ;.
N31047
g1151446
EST.
human clone=265309 primer=77 library=Soares melanocyte 2NDHM
vector=pr773D (Pharmacia) with a modified polylinker host-DH10B
(ampicillin resistant) Rsite1-Not I Rsite2-Eco RI Male. 1st strand
cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGGCGCGCAGTGTGTGTGTGTGTGTGTGT-3'],
double-stranded cDNA was size selected, ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I and Eco
RI sites of a modified pr773 vector (Pharmacia). Library
constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal
foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P.
Albino.
ORGANISM      Homo sapiens
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Homidae; Homo.
REFERENCE     1 (bases 1 to 490)
AUTHORS       Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
               Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
               Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
               Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
               Wilson,R.
TITLE         The WashU-Merck EST Project
JOURNAL       Unpublished (1995)
COMMENT       Contact: Wilson RK
               WashU-Merck EST Project
               Washington University School of Medicine
               4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
               Tel: 314 286 1800
               Fax: 314 286 1810
               Email: est@watson.wustl.edu
               High quality sequence stops: 385
               Source: IMAGE Consortium, LLNL
               This clone is available royalty-free through LLNL; contact the
               IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES      source
               1..490
               /organism="Homo sapiens"
               /clone="265309"
               <1..>490
BASE COUNT    131 a 107 c 111 g 138 t 3 others
ORIGIN
Query Match 7.7%; Score 95; DB 92; Length 490;
Best Local Similarity 67.5%; Pred. No. 2.29e-89;
Matches 183; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
Db 107 gaagggcaggtgtgtcttctattcagggcactcttcttccatgtactgcagctgttt 166
QY 478 GAGGCCAGGTGTCTTCTACTCGGACACTCTTCTTGGGATGTACTGCATGGTCTC 537
Db 167 gtggcactttatcttcaagccagatgaagggagactgggcaagactcttaccnccaca 226
QY 538 TTGGCGTGTATGTATGACAGGACAGACTCTGTGTGAAGTGGGACGCTGCTGCGACCCACA 597
Db 227 ctgcaatttggtctgttgcgtatcattatgtatggcctttctcagttctcagttat 286
QY 598 GTCCAGTCTTCTCTGGTGGCCCTTGGCCCTCTACGTGGGCTACACCCGGGTCTGTATAC 657
Db 287 aaacaccactggagcagatgtgttgactggactcattcaggagagctcgtgttgcaatatta 346
QY 658 AAACACCACCTGGAGCGATGTCTTGTGGCTCTCTGCGAGGGGACACTGGTGGCTGCCCTC 717
Db 347 gttgctgtatgtatcggattcttccaaag 377
QY 718 ACTGTCTGCTACATCTCAGACTTCTTCAAAG 748

```

```

RESULT 14
LOCUS R63796 446 bp mRNA EST 26-MAY-1995
DEFINITION Y115H04.r1 Homo sapiens cDNA clone 139351 5'.
ACCESSION R63796
NID 9835675
KEYWORDS EST.
SOURCE human clone-139351 library=Soares placenta NB2HP vector-pt7T3D
 (Pharmacia) with a modified polylinker host=DH10B (ampicillin
 resistant), primer=M13RP1 Rsite1-Not I Rsite2-Eco RI Female placenta
 obtained at birth (full term). 1st strand cDNA was primed with a
 Not I - oligo(dT) primer [5',
 AACTGGAAGAAATCGGGCGGAGGAAATTTTTTTTTTTT 3'], double-stranded
 cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not
 I and cloned into the Not I and Eco RI sites of the modified pT7T3
 vector. Library went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM Homo sapiens
Eukaryote; Metazoa; Eumetazoa; Bilateria; Chordata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Chondata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 446)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevisakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 153
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES
source
Location/Qualifiers
1..446
/organism="Homo sapiens"
/clone="139351"
BASE COUNT 84 a 122 c 110 g 124 t 6 others
ORIGIN
Query Match 6.1%; Score 75; DB 29; Length 446;
Best Local Similarity 60.8%; Pred. No. 3,28e-62;
Matches 205; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

Db 8 tgggctgttctcttggctgtgctcaagccagcttttcacagacattgccaaagtgt 67
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 299 TGGGACCTTCTGTTTGGGCTGCGCGTAGCCAGCTCTCTGACAGACCTGCGCAAGTACA 358
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 68 coataaggcgctgcgtcactcttggatgtctgcaacctgatttcagccagatca 127
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 359 TGATTGGCGCTGTAAGCCCACTTCTTAGCGCTGTGGACCCGACCTGAGCGGGTCA 418
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 128 actgctctgaagggtacattcagaactacagatgcagaggtgatgacagcaagtcagg 187
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 419 ACTGCTCGTCTATGTGCAGCTGGAGAGGTGTGCAGGGGAAACCTCTGATGTCACCG 478
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 188 naggcaggnagtccttcttcttgccatgtcttccttcttcactatgcattgtatt 247
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 479 AGGCCAGATTCTTCTACTCGGGACACTCTTCTTTGGATGTAATGCTGTTCTTCT 538
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 248 tgggtctactctgagccctcttcatttggcgagagcgccgtctccgcccctcc 307
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 539 TGGCGCTGATGTGCAGGACAGACTCTGTTGGAAGTGGGCGAGCGGTCTGCGACCCACAG 598
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 308 tycagttcaccttgatgatggccttctacacggg 344

```

```

Qy 599 TCAGTTCCTCCTGGTGGCTTTTGCCTTACGTGG 635
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
LOCUS W04968 472 bp mRNA EST 23-APR-1996
DEFINITION za43309.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
295360 5' similar to WP:728D9.3 CE02068 ;
ACCESSION W04968
NID g1277708
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Eumetazoa; Bilateria; Chordata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Chondata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 472)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevisakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: mob.REGA+ET
High quality sequence stop: 339.
FEATURES
source
Location/Qualifiers
1..472
/organism="Homo sapiens"
/notes="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACTGGAGAAATTAATAAGATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
/clone="295360"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
BASE COUNT 120 a 102 c 98 g 152 t
ORIGIN
Query Match 5.8%; Score 72; DB 114; Length 472;
Best Local Similarity 62.1%; Pred. No. 3.14e-58;
Matches 211; Conservative 0; Mismatches 127; Indels 2; Gaps 2;

Db 106 aataactacatgacactattacaagccattggaacctttttatttggcgactgt 165
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 268 AACAACTACGTGGCTGTGTATACAGGTGCTGGGACCTTCTCTGTTGGGCTGCCGTG 327
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 166 agtcagtcctctgactgacattgccaagtattcaataggcagactggcgccactcttt 225
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 328 AGCCAGTCTCTGACAGACCTG-GCCAAAGTACATGATTGGCGCTCTGAAGCCCACTTCT 386
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 226 ggaattgtgatccagattggtcaaaaaatacaactgcaagcgatgggtacattgaata 285
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 387 AGCCGTCTGCAACCCCGGCTCAACTGCTCGGTCTATGTGCAAGCTGG-AGA 445
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 286 acatatgtcggaggaatgcagaaagagtgtaaggagggcaggtgttccttctattcaggcc 345

```

```

QY 446 AGGTGTGCAGGGGAAACCCCTGCTGATGTCACCGAGGCCAGGTTGCTTCTACTCGGGAC 505
Db 346 actcttcgtttccatgtagctgcatgctgtttgtggcactttatcttcaagccaggatga 405
QY 506 ACTCTTCCTTTGGGATGTACTGCTGCTGCTTGGCGCTGATGTGCAGGCACGACTCT 565
Db 406 agggagactgggcaagactcttacgccccacactgcaatt 445
QY 566 GTTGGAAAGTGGGCACGGCTGCTGGACCCACACAGTCCAGTT 605

```

Search completed: Thu Nov 6 12:58:09 1997
Job time : 698 secs.

W A P E S R E L E A S E
***** (TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by Intelligenetics, Inc.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Thu Nov 6 12:58:35 1997; MasPar time 336.04 Seconds
Tabular output not generated. 1007.529 Million cell updates/sec

Title: >US-08-842-827-7
Description: (1-1232) from US08842827.seq
Perfect Score: 1232
N.A. Sequence: 1 ACCATGCGAGCGAGTGGGT.....CAAAAAAAAAAAAAAAAAA 1232
Comp: TGGTAGCGTGCCTCCACCA.....GTTTTTTTTTTTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 359085 seqs, 137405154 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: EST-STS-THREE
1:EST199 2:EST200 3:EST201 4:EST202 5:EST203 6:EST204
7:EST205 8:EST206 9:EST207 10:EST208 11:EST209 12:EST210
13:EST211 14:EST212 15:EST213 16:EST214 17:EST215
18:EST216 19:EST217 20:EST218 21:EST219 22:EST220
23:EST221 24:EST222 25:EST223 26:EST224 27:EST225
28:EST226 29:EST227 30:EST228 31:EST229 32:EST230
33:EST231 34:EST232 35:EST233 36:STS1 37:STS2 38:STS3
39:STS4 40:STS5 41:STS6 42:STS7 43:STS8 44:STS9 45:STS10
46:STS11 47:STS12 48:STS13

Database: EST-STS-FOUR
49:gnEST1 50:gnEST2 51:gnEST3 52:gnEST4 53:gnEST5
54:gnEST6 55:gnEST7 56:gnEST8 57:gnEST9 58:gnEST10
59:gnEST11 60:gnEST12 61:gnEST13 62:gnEST14 63:gnEST15
64:gnEST16 65:gnEST17 66:gnEST18 67:gnEST19 68:gnEST20
69:gnEST21 70:gnEST22 71:gnEST23 72:gnEST24 73:gnEST25
74:gnSTS 75:gnEST1 76:gnEST2 77:gnEST3 78:gnEST4
79:gnEST5 80:gnEST6 81:gnEST7 82:gnEST8 83:gnEST9
84:gnEST10 85:gnEST11 86:gnEST12 87:gnEST13 88:gnEST14
89:gnEST15 90:gnEST16 91:gnEST17 92:gnEST18 93:gnEST19
94:gnEST20 95:gnEST21 96:gnEST22 97:gnEST23 98:gnEST24
99:gnEST25 100:gnEST26 101:gnEST27 102:gnEST28
103:gnEST29 104:gnEST30 105:gnEST31 106:gnSTS 107:ueEST1
108:ueEST2

Statistics: Mean 11.492; Variance 3.475; scale 3.307

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result Query

SUMMARIES

No.	Score	Match	Length	DB	ID	Description	Pred. No.
c 1	458	37.2	491	25	N54789	yv31b01.sl Soares fet	0.00e+00
2	395	32.1	478	26	N75714	yv31b01.rl Soares fet	0.00e+00
3	66	5.4	385	33	AA107326	ml95b10.rl Stratagene	2.53e-46
4	61	5.0	337	27	W39816	309 Mouse VM cDNA lib	2.96e-39
5	60	4.9	402	27	W39811	307 Mouse VM cDNA lib	4.40e-39
6	59	4.8	564	1	W30942	zc64f09.rl Soares fet	6.47e-38
7	59	4.8	564	88	HS942332	zc64f09.rl Soares fet	6.47e-38
8	56	4.5	242	27	W39815	308 Mouse VM cDNA lib	1.94e-34
9	47	3.8	87	33	AA106892	ml85a07.rl Stratagene	2.62e-24
c 10	42	3.4	313	91	MM1155742	mx78g01.rl Soares mou	6.44e-19
c 11	42	3.4	313	66	AA237956	mx78g01.rl Soares mou	6.44e-19
12	41	3.3	415	87	HS605336	zb63f08.rl Soares fet	7.27e-18
13	34	2.8	914	34	AA141702	CK02248.contig Drosop	8.73e-11
c 14	32	2.6	422	34	AA142082	CK02016.3prime Drosop	7.12e-09
c 15	30	2.4	446	12	AA181403	zp52f10.sl Stratagene	5.05e-07
c 16	28	2.3	228	87	HS550335	zb91e04.sl Soares par	3.05e-05
17	28	2.3	236	81	HS1146241	zp52f10.rl Stratagene	3.05e-05
18	28	2.3	236	12	AA182434	zp52f10.rl Stratagene	3.05e-05
c 19	28	2.3	263	77	AT6391	2647 Arabidopsis thal	3.05e-05
c 20	28	2.3	343	63	AA225572	nc08c05.sl NCI CGAP P	3.05e-05
c 21	28	2.3	343	81	HS1144996	nc08c05.sl NCI CGAP P	3.05e-05
c 22	28	2.3	435	14	AA187745	zp72g10.sl Stratagene	3.05e-05
c 23	28	2.3	495	11	AA176159	zp23h04.sl Stratagene	3.05e-05
24	27	2.2	100	6	AA120007	mn15h06.rl Beddington	2.22e-04
25	27	2.2	252	99	MMAA31554	mx13f08.rl Soares mou	2.22e-04
26	27	2.2	252	64	AA2311554	mx13f08.rl Soares mou	2.22e-04
27	27	2.2	260	101	MMAA60919	mx77h08.rl Soares mou	2.22e-04
28	27	2.2	336	91	MM1156654	mx92d02.rl Soares mou	2.22e-04
29	27	2.2	336	67	AA238805	mx92d02.rl Soares mou	2.22e-04
c 30	27	2.2	347	39	G09147	human STS CHLC.GATA2B	2.22e-04
31	27	2.2	351	101	MMAA60235	va95d02.rl Soares mou	2.22e-04
32	27	2.2	385	78	AT9619	10739 arabidopsis tha	2.22e-04
c 33	27	2.2	395	3	AA112233	zm64b11.sl Stratagene	2.22e-04
c 34	27	2.2	403	87	HS733324	zc64f09.sl Soares fet	2.22e-04
c 35	27	2.2	403	1	W02733	zc64f09.sl Soares fet	2.22e-04
c 36	27	2.2	405	6	AA121078	zm22b08.rl Stratagene	2.22e-04
c 37	27	2.2	432	99	MMAA37207	mx18e02.rl Soares mou	2.22e-04
38	27	2.2	432	66	AA237207	mx18e02.rl Soares mou	2.22e-04
c 39	27	2.2	435	33	AA105273	ml82h08.rl Stratagene	2.22e-04
40	26	2.1	167	102	MMAA74695	mt13b09.rl Soares mou	1.54e-03
c 41	26	2.1	210	84	HS1154328	zs05e01.sl Soares NbH	1.54e-03
42	26	2.1	335	67	AA238825	mx92g02.rl Soares mou	1.54e-03
43	26	2.1	335	91	MM1156666	mx92g02.rl Soares mou	1.54e-03
44	26	2.1	343	5	AA117600	mp65a04.rl Soares 2Nb	1.54e-03
c 45	26	2.1	461	30	AA063613	ze87f05.sl Soares fet	1.54e-03

ALIGNMENTS

RESULT	1	N54789	491 bp	MRNA	EST	28-JAN-1997
LOCUS	yv31b01.sl Soares fetal liver spleen	INFLS	Homo sapiens	CDNA	clone	
DEFINITION	244297 3'					
ACCESSION	N54789					
NID	g1196109					
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.					
AUTHORS	1 (bases 1 to 491)					
	Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.					
TITLE	WashU-Merck EST Project					
JOURNAL	Unpublished (1995)					
COMMENT	Contact: Wilson RK WashU-Merck EST Project Washington University School of Medicine					


```

Db 301 tggccaagtacatgattggcgctgagggcccaactttctagccgtctcgcaanccgact 360
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 347 TGGCAAGTACATGATTGGCGCTGAAGCCCAACTTCTAGCGCTGCGACCCGACT 406
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 361 ggagccgggttaa-tgctcggtctatnctggaaacttggaaaagtgttgaagggaacct 419
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 407 GGAGCCGGGTCAACTGCTCGGTCTATGTGCACT-GGAGAAAGGTGCGAGGGAACCCCT 465
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 420 gctatgtaaacaggcgaagt-gctttcaa-tcggaataatctcttt 467
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 466 CTTGATGTACAGGAGCCAGGTGCTTCTTACTCGGACACTCTTCCTT 515
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
LOCUS AA107326 385 bp mRNA EST 04-FEB-1997
DEFINITION ml95b10_r1 Stratagene mouse kidney (#937315) Mus musculus cDNA
clone 519739 5' similar to TR:G1161100 G1161100 HYDROGEN
PEROXIDE-INDUCIBLE PROTEIN ;
ACCESSION AA107326
NID 91659299
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;
Mus.
REFERENCE 1 (bases 1 to 385)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LILN ; contact the
IMAGE Consortium (info@image.liln.gov) for further information.
MGI:313587
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 375.
FEATURES
source
location/Qualifiers
1..385
/organism="Mus musculus"
/strain="C57/B16"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dr. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor
sequence: 5' GAATTCGGCAGGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTCTTTTCTTTT 3'"
/clone="519739"
/clone_lib="Stratagene mouse kidney (#937315)"
/sex="females"
/dev_stage="4 weeks"
/lab_host="SOLR (kanamycin resistant)"
<1..>385
BASE COUNT 95 a 84 c 89 g 117 t
ORIGIN
Query Match 5.4%; Score 66; DB 33; Length 385;
Best Local Similarity 62.5%; Pred. No. 3 53e-46;
Matches 178; Conservative 0; Mismatches 106; Indels 1; Gaps 1;

Db 69 ctacatagccaccattacaagcgctggagccctttttgttcgagtcacagtagtca 128
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 273 CTACGTGGCTGTATACAGGTGCTGGGACCTTCTCTTTGGGGTTCGCGTGAGCCA 332
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 129 gtccttgactgacatcgctaaagtactatagcagtttgcgcga-cttcttggtcat 187
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 333 GTCTCTGACAGACCTGGCCAAAGTACATGATTGGCGCTGAAGCCCAACTTCTTAGCCGT 392
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 188 ctgtaaccacagactgggtcaaaaaaactgaactgcagtgatggctatattgaggactacatg 247
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 393 CTGGACCCCGCACTGGAGCCGGGTCAACTGCTCGGTCTATGTGAGCTGGAGAGGTGTG 452
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 248 tcaagggaatgaagagaaagtcaaggaggcgaggttcttcttcttactcgggacactcttc 307
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 453 CAGGGGAACCCCTGCTGATGTACACGAGGCCAGGTGTCTTCTACTCGGACACTCTTTC 512
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 308 attctatgtactgcatgctgtttgttcgcaacttattcttcaagc 352
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 513 CTTTGGGATGTACTGCATGGTGTCTTGGCGCTGTATGTGAGGC 557
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
LOCUS W39816 337 bp mRNA EST 05-FEB-1997
DEFINITION 309 Mouse VM cDNA library Mus musculus cDNA clone spmp22 1.25.
ACCESSION W39816
NID 91816979
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;
Mus.
REFERENCE 1 (bases 1 to 337)
AUTHORS Stewart,G.J., Savioz,A. and Davies,R.W.
TITLE Sequence analysis of 497 mouse brain ESTs expressed in the
substantia nigra
JOURNAL Genomics 39, 147-153 (1997)
COMMENT
Contact: Davies,R.W.
Robertson Laboratory of Biotechnology
Institute of Biomedical and Life Sciences, Division of Molecular
Genetics
Robertson Building, 54 Dumbarton Road, Glasgow G11 6NU
Tel: 44 141 330 5102
Fax: 44 141 330 5102/4878
Email: gbga21@udcf.gla.ac.uk
Seq primer: T7/T3alpha
High quality sequence stop: 337.
FEATURES
source
location/Qualifiers
1..337
/organism="Mus musculus"
/notes="Organ: brain; Vector: pSPORT1; Mouse Ventral
Midbrain directional cDNA library in pSPORT1. The library
was created by subtractive hybridisation using VM mRNA as
the target nucleic acid population and total cerebellar ss
cDNA as the driver nucleic acid population. The clones are
3' directed."
/clone="spmp22 1.25"
/clone_lib="Mouse VM cDNA library"
/tissue_type="ventral midbrain"
<1..>337
BASE COUNT 59 a 112 c 83 g 83 t
ORIGIN
Query Match 5.0%; Score 61; DB 27; Length 337;
Best Local Similarity 62.3%; Pred. No. 2.96e-40;
Matches 154; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Db 1 cgggcccagcctctctccatgttcactatgcgtatctgtgtctctaccctcaggccc 60
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 500 CGGACACTCTTCTCTTGGGATGATCTCATCGATGGTGTCTTGGCGCTGTATGTGAGGCAC 559
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 61 gcttcactgcgcggggcccgccgtctccgcctccctcctgcagttcactttgttcattga 120
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 560 GACTCTGTGGAAGTGGCGACGGCTGCTGCGACCCACACAGTCCAGTCTTCTCTGTGGGCT 619
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```

||||| 328 ACCAGTCTTGACAGACCTGGCCAAAGTACATGATGGCGCTCTGAAGCCCACTTCT- 386
||||| 311 aagtgtctgcaacctgatttcagcagatcaactgctctgaaggctacattcagaacta 370
||||| 387 AGCGTCTGCGACCCGACTGGAGCCGGGTCACTGCTCGTCTATGTGCGACCTGGAGAA 446
||||| 371 cagatgcagagtgatgacagcaagtcacagangccagggaagtcctctcttgccca 430
||||| 447 GGTGTGACAGGGAACCCCTGCTGATGTCACGAGGCCAGGTTGCTTTCTACTCGGGACA 506
||||| 431 tgcctctctccatgacatgctgaagtattggtctatantctgagcccgcttca 490
||||| 507 CTCCTCCCTTGGGAGTACTGCATGGTGTCTT-GGCGCTGATGTGCGAGCAGCACTCT 565
||||| 491 cttgcgagagagc-cgncctgctnccggccctcctcgactt 529
||||| 566 GTTGAAGTGGGACGGCTGCTGGACCCACAGTCCAGTT 605

RESULT 7
ID HS942332 standard; RNA; EST; 564 BP.
AC W30942;
NI Q1311934
DT 13-MAR-1996 (Rel. 47, Created)
DE 07-MAR-1997 (Rel. 51, Last updated, Version 2)
DE zc64f09.r1 Soares fetal heart NBHL19W Homo sapiens cDNA clone
DE 327113 5' similar to WP:28D9.3 CE02068 ;.
KW EST.
OS Homo sapiens (human)
OC Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
OC Vertebrata; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP 1-564
RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,
RA Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,
RA Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F.,
RA Trevaaskis E., Waterston R., Williamson A., Wohlmann P., Wilson R.;
RA "The WashU-Merck EST Project";
RL Unpublished.
CC Contact: Wilson RK WashU-Merck EST Project Washington University
CC School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,
CC MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:
CC est@wustl.wustl.edu This clone is available royalty-free through
CC LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for
CC further information. Insert Length: 774 Std Error: 0.00 Seq primer:
CC mob.REGA+ET High quality sequence stop: 336.
FH Key Location/Qualifiers
FH source
FH 1..564
FH /organism="Homo sapiens"
FH /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
FH modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
FH strand cDNA was primed with a Not I - oligo(dT) primer [5'
FH TGTACCAATCTGAAGTGGAGCGCGCGATCTTTTTTTTTTTT 3'],
FH double-stranded cDNA was size selected, ligated to Eco RI
FH adapters (Pharmacia), digested with Not I and cloned into
FH the Not I and Eco RI sites of a modified pT7T3 vector
FH (Pharmacia). Library went through one round of
FH normalization to a Cot = 5. Library constructed by M.Fatim
FH a
FH Bonaldo. This library was constructed from the same fetus
FH as the fetal lung library, Soares fetal lung NBHL19W."
FH /clone="327113"
FH /clone_lib="Soares fetal heart NBHL19W"
FH /sex="unknown"
FH /dev_stage="19 weeks"
FH /lab_host="DH10B (ampicillin resistant)"
FH <1..>564
FH mRNA
SQ Sequence 564 BP; 128 A; 155 C; 139 G; 134 T; 8 other;

Query Match 4.8%; Score 59; DB 88; Length 564;
Best Local Similarity 60.9%; Pred. No. 6.47e-38;
Matches 149; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

```

```

Matches 207; Conservative 0; Mismatches 129; Indels 4; Gaps 4;
Db 192 aacccctacgtggcagcactctataaagtggtgctctcttctgtgca-tc 250
||||| 268 AACAACTACGTGGCTGCTGTATACAAGGTGCTGGGACCTTCTGTTGGGCTGCCGTG 327
||||| 251 agccagctcttcacagacattgccaagtgtccataaggcgctgcctcactcttgg 310
||||| 328 AGCCAGTCTCTGACAGACCTGGCCAAAGTACATGATGGCGCTCTGAAGCCCACTTCT- 386
||||| 311 aagtgtctgcaacctgatttcagcagatcaactgctctgaaggctacattcagaacta 370
||||| 387 AGCGTCTGCGACCCGACTGGAGCCGGGTCACTGCTCGTCTATGTGCGACCTGGAGAA 446
||||| 371 cagatgcagagtgatgacagcaagtcacagangccagggaagtcctctcttgccca 430
||||| 447 GGTGTGACAGGGAACCCCTGCTGATGTCACGAGGCCAGGTTGCTTTCTACTCGGGACA 506
||||| 431 tgcctctctccatgacatgctgaagtattggtctatantctgagcccgcttca 490
||||| 507 CTCCTCCCTTGGGATGATGCATGGTGTCTT-GGCGCTGATGTGCGAGCAGCACTCT 565
||||| 491 cttgcgagagagc-cgncctgctnccggccctcctcgactt 529
||||| 566 GTTGAAGTGGGACGGCTGCTGGACCCACAGTCCAGTT 605

RESULT 8
LOCUS W39815 242 bp mRNA EST 05-FEB-1997
DEFINITION 308 Mouse VM cDNA library Mus musculus cDNA clone spmp22 1.23.
ACCESSION W39815
NID gi1816978
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Mus.
REFERENCE 1 (bases 1 to 242)
AUTHORS Stewart,G.J., Savioz,A. and Davies,R.W.
TITLE Sequence analysis of 497 mouse brain ESTs expressed in the
SUBMITTER Genomics 39, 147-153 (1997)
JOURNAL
COMMENT Contact: Davies,R.W.
Robertson Laboratory of Biotechnology
Institute of Biomedical and Life Sciences, Division of Molecular
Genetics
Robertson Building, 54 Dumbarton Road, Glasgow G11 6NU
Tel: 44 141 330 5102
Fax: 44 141 330 5102/4878
Email: gbg21@udcf.gla.ac.uk
Seq primer: T7/T3alpha
High quality sequence stop: 2.
FEATURES
source
1..242
/organism="Mus musculus"
/note="Organ: brain; Vector: pSPORT1; Mouse Ventral
Midbrain directional cDNA library in pSPORT1. The library
was created by subtractive hybridisation using VM mRNA as
the target nucleic acid population and total cerebellar ss
cDNA as the driver nucleic acid population. The clones are
3' directed."
/clone="spmp22 1.23"
/clone_lib="Mouse VM cDNA library"
/tissue_type="ventral midbrain"
<1..>242
BASE COUNT 34 a 82 c 60 g 66 t
ORIGIN

Query Match 4.5%; Score 56; DB 27; Length 242;
Best Local Similarity 61.6%; Pred. No. 1.94e-34;
Matches 149; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

```


Search completed: Thu Nov 6 13:05:06 1997
Job time : 391 secs.

This Page Blank (uspto)

US5508384-A.
16-APR-1996.
PF 10-SEP-1992; 943236.
PR 10-SEP-1992; US-943236.
PR 09-SEP-1993; US-118270.
PA (UYNV) UNIV NEW YORK STATE.
PI Murphy RB, Schuster DI;
PT WPI; 96-208785/21.
PT New dopamine receptor peptide - useful as antipsychotic agent, e.g.
PT for treating schizophrenia
PS Disclosure; Column 49-52; 184pp; English.
SC Proteins W02657-W02720 represent a range of G-protein coupled receptor
CC (GPR) proteins selected from cAMP, adenosine, muscarinic acetylcholine,
CC adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
CC odorant, cytomagaloviral and other GPR proteins. The receptor proteins
CC were used to design polypeptides, pref. based on the transmembrane
CC domains, for use in G-protein coupled receptor ligand binding assays.
CC The polypeptide fragments retain biological activity such as binding a
CC GPR ligand or modulating GPR ligand binding to a GPR (see W02747-W02999
CC for examples of polypeptide fragments). The polypeptide fragments can
CC be used in compositions for treating subjects suffering from a pathology
CC related to a GPR abnormality e.g. a psychotic disorder such as
CC schizophrenia.
SQ Sequence 317 AA;

```

Query Match      4.68; Score 97; DB 19; Length 317;
Best Local Similarity 25.99; Pred. No. 2.Ole+01;
Matches 15; Conservative 23; Indels 3; Gaps 3;

Db 66 vvedlyvsna-svmnlllisfr-yfcvktptypvkrttkmagnmiaawlsfil 121
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 12 VLC-LLVASLPEAILTVNAPYKRGVCGDSSIRYPYRPTTIHGLMAGVTITATVIL 68
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 3

ID R48689 standard: Protein; 317 AA.

AC R48689;

DE 05-JUN-1996 (first entry)

DE G-protein coupled human m2 muscarinic acetylcholine receptor protein.

DE G-protein coupled receptor; ligand binding assay; transmembrane domain;

KW psychotic disorder; schizophrenia; dopamine; cAMP; adenosine; thrombin;

KW muscarinic acetylcholine; adrenergic; endothelin; bombesin; endocrine;

KW rhodopsin; opsin; odorant; cytomegalovirus.

OS Homo sapiens.

PN WO9405695-A1.

PN 17-MAR-1994.

PD 09-SEP-1993; U08528.

PR 10-SEP-1992; US-943236.

PA (UYNV) UNIV NEW YORK STATE.

PI Murphy RS, Schuster DI;

DR WPI; 94-101120/12.

PT Polypeptides of G-coupled receptor proteins (GPRs) - useful for

PT binding GPR ligands or modulating GPR binding

PS Disclosure; Page 61: 160pp; English.

CC Proteins R48685-R48758 represent a range of G-protein coupled receptor

CC proteins selected from cAMP, adenosine, endothelin, bombesin, opsin,

CC adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin,

CC odorant, cytomegaloviral and other G-protein coupled receptors. The

CC receptor proteins were used to design polypeptides, pref. based on the

CC transmembrane domains, for use in G-protein coupled receptor ligand

CC binding assays. The polypeptide fragments retain biological activity

CC such as binding a GPR ligand or modulating GPR ligand binding to a GPR

CC (see R48759-R48758, R50569-R50807 and R80189-R83195 for examples of

CC polypeptide fragments). The polypeptide fragments can be used in

CC compositions for treating subjects suffering from a pathology related to

CC a GPR abnormality e.g. a psychotic disorder such as schizophrenia.

SQ Sequence 317 AA;

Query Match 4.6%; Score 97; DB 16; Length 317;
Best Local Similarity 25.9%; Pred. No. 2.01e+01;
Matches 15; Conservative 17; Mismatches 23; Indels 3; Gaps 3;
Db 66 vcdlyvvsna-svynllllisfdr-yfcvckbltpypvkrtrtkmagnmiaaawlsfil 121

```

QY      12 VLC-LLVASLFFAILTLVNAPKRGFYCGDSDSIRYPDPOTITHGLMAGVYTATVIL 68
        | : | | | | : : : : | : | : | : | : | : | : | : | : | : |
RESULT      4
ID       R05233 standard; protein; 172 AA.
AC       R05233;
DT       03-AUG-1990 (first entry)
DE       Human lymphotoxin (HL) encoded by new synthetic lymphotoxin gene
KW       Human lymphotoxin (HL).
PN       J02000447-A.
PD       05-JAN-1990.
PF       27-OCT-1987; 271513.
PR       27-OCT-1987; JP-271513, JP-287035.
PA       (SANY) Sankyo Kk.
PI       WPI; 90-047988/07.
DR       DR N-PSDB; Q93289
PT       Human lymphotoxin synthetic gene -
PT altered to leave only one restriction enzyme recognition site
PS Disclosure; Fig 1; 13pp; Japanese.
CC It is the same amino acid sequence as natural HL. Its encoding ge
CC genetically engineered to make modifications of HL easier and impr
CC lymphotoxins can be prepd.
SQ Sequence 172 AA:

```

```

Query Match      4.5%  Score 95;  DB 1;  Length 172;
Best Local Similarity 26.2%;  Pred. No. 2.73e+01;
Matches 11;  Conservative 16;  Mismatches 15;  Indels 0;  Gaps 0;

Db 47 rantdrflgdgfslnsncsllyvptsglyfyvsqvfvrkays 88
   | | | | | : : : : : | | | : : | | : : | | :
Qy 48 RPDITITGLMGAGVITATVILNVSAGEAYLVYTORLXSRSDFN 89

```

RESULT 5
ID R06584 standard; protein; 2020 AA.
AC R06584;
DT 10-JAN-1991 (first entry)
DE Cardiac sodium channel.
KW Rat; arrhythmia.
OS Rattus rattus.
PN W09009391-A.
PD 23-AUG-1990.
PF 09-FEB-1990; U000768.
PR 13-FEB-1989; US-310330.
PA (ARCH-) ARCH DEV CORP.
PI Rogart RB:
DR WPI; 90-275095/36.
DR N-PSDB; Q05831.
PT New rat cardiac sodium channel proteins - and associated DNA
PT sequences, polypeptides and peptides associated with
PT proteins, useful as antiarrhythmic and cardiotoxic drugs.
PS Disclosure; Fig 2: 65pp; English.
CC The sequence deduced from cDNA derived from 3 overlapping clones,
CC PH3-1, pRH4-23, and pRH14-31. (Deposited as ATCC 67885, 67886, in
CC and 67887 resp.) The clones were isolated from a cDNA library in
CC the lambda Zap vector prep. from mRNA obtd. from newborn rat
CC hearts using rat brain II cDNA probe. The protein has diagnostic,
CC therapeutic, and prognostic applications.
CC Sequence 2020 AA;

```
Query Match      4.5%: Score 95; DB 2; Length 2020;
Best Local Similarity 7.7%; Pred. No. 2.73e+01;
Matches 13; Conservative 15; Mismatches 16; Indels 3; Gaps 3;

Db 352 gytsdfsfawflalfmlmtgdcwerlyqqtllrsagklymffmlvl 398
|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
Oy 167 GHSGFGMYCMVFLALY-VQARLCW-K-WARLLRPTVQFFLVAVALYV 210

RESULT 6
ID P50101 standard; Protein: 154 AA.
```


AC P50101;
 DE 30-OCT-1991 (first entry)
 DE Sequence of a pure polypeptide having tumour necrosis factor
 DE (TNF) activity.
 KW Antiviral agent; lymphokine; antitumour.
 OS Oryctolagus cuniculus.
 PN EP-148311-A.
 PD 17-JUL-1985.
 PF 07-MAY-1984; 105149.
 PR 26-DEC-1983; JP-251817.
 PR 08-FEB-1984; JP-019850.
 PR 18-APR-1984; JP-076584.
 PA (ASAH.) ASAH KASEI KOGYO.
 PI Itoh H;
 DR N-PSDB; 85-172835/29.
 DR N-PSDB; N50113.
 PT New polypeptide(s) having 154 aminoacid sequence - useful for
 PT inducing necrosis of tumours and as antivirals
 PS Claim 1; Page 54; 65pp; English.
 CC The TNFs of the invention are able to induce necrosis of tumours
 CC without toxic effects on the normal living body tissues. They are
 CC also antivirals. Dose is 50-100 x 1,000,000 units daily for an adult.
 CC The DNA is ligated to a replicable expression vehicle to give a
 CC recombinant DNA. This is used to transform a microorganism,
 CC transformants are selected and incubated to produce TNF.
 CC Specifically claimed are plasmids pTNF-lac-1 and pTHF-lacUV5-1 and
 CC E.coli K-12 strain transformed with them. The DNA used was
 CC identified and isolated after studies on rabbit TNF.
 SQ Sequence 154 AA;
 Query Match 4.4%; Score 94; DB 3; Length 154;
 Best Local Similarity 29.8%; Pred. No. 3.18e+01;
 Matches 14; Conservative 14; Mismatches 18; Indels 1; Gaps 1;
 Db 31 anallangmkldnqlvvpadgyllysqvlfsgggcrsyvllthtv 77
 QY 53 THGLMA-GVTITATVILVSAGEAYLVYTDRLYSRDFNNVAAVYKV 98
 RESULT 7
 ID R05175 standard; protein; 154 AA.
 AC R05175;
 DT 18-JUL-1990 (first entry)
 DE Tumoricidal polypeptide.
 KW Tumoricide; Tumour necrosis factor.
 PN J02019398-A.
 PD 23-JAN-1990.
 PF 06-JUL-1988; 166913.
 PR 06-JUL-1988; JP-166913.
 PA (ASAH) Asahi Chemical Ind KK.
 DR WPI; 90-064642/09.
 DR N-PSDB; Q02035.
 PT Tumoricidal polypeptide(s) - in which DNA to code the polypeptide(s) is
 PT constructed from human-derived TNF-coding DNA and rabbit-derived
 PT TNF-coding DNA.
 PS Disclosure; Page 3; 11pp; Japanese.
 CC The sequence is derived from tumour necrosis factor DNA.
 CC See also R05173-76 and R05189-90.
 SQ Sequence 154 AA;
 Query Match 4.4%; Score 94; DB 2; Length 154;
 Best Local Similarity 29.8%; Pred. No. 3.18e+01;
 Matches 14; Conservative 14; Mismatches 18; Indels 1; Gaps 1;
 Db 31 anallangmkldnqlvvpadgyllysqvlfsgggcrsyvllthtv 77
 QY 53 THGLMA-GVTITATVILVSAGEAYLVYTDRLYSRDFNNVAAVYKV 98
 RESULT 8
 ID P50138 standard; protein; 154 AA.
 AC P50138;
 DT 01-JAN-1980 (first entry)

DE Rabbit tumor necrosis factor.
 KW Tumor necrosis factor; lymphokine; antitumor; ss.
 OS Oryctolagus coturnix.
 PN EP-148026-A.
 PD 26-JUN-1985.
 PF 27-NOV-1984; 114325.
 PR 02-DEC-1983; JP-228790.
 PA (DAIN) DAINIPPON PHARM KK.
 PI Yamda M, Furutani Y, Notake M;
 DR WPI; 85-154142/26.
 DR N-PSDB; N50136.
 PT New DNA encoding rabbit tumour necrosis factor - useful in
 PT expression of the factor in Escherichia coli for anti:tumour use
 PS Claim 2; Page 39; 64pp; English.
 CC The rabbit TNF can be obtained in large amounts by recombinant DNA
 CC techniques. DNA encoding rabbit TNF is obtained by cultivation of
 CC rabbit macrophages with inducers. A fraction containing TNF mRNA is
 CC separated from the macrophages and used for the preparation of a
 CC cDNA library. The TNF may be used as an anti:tumor agent. See also
 CC N50197 and P50139.
 SQ Sequence 154 AA;
 Query Match 4.4%; Score 94; DB 3; Length 154;
 Best Local Similarity 29.8%; Pred. No. 3.18e+01;
 Matches 14; Conservative 14; Mismatches 18; Indels 1; Gaps 1;
 Db 31 anallangmkldnqlvvpadgyllysqvlfsgggcrsyvllthtv 77
 QY 53 THGLMA-GVTITATVILVSAGEAYLVYTDRLYSRDFNNVAAVYKV 98
 RESULT 9
 ID R05190 standard; protein; 154 AA.
 AC R05190;
 DT 18-JUL-1990 (first entry)
 DE Tumoricidal polypeptide.
 KW Tumoricide; Tumour necrosis factor.
 PN J02019398-A.
 PD 23-JAN-1990.
 PF 06-JUL-1988; 166913.
 PR 06-JUL-1988; JP-166913.
 PA (ASAH) Asahi Chemical Ind KK.
 DR WPI; 90-064642/09.
 DR N-PSDB; Q02045.
 PT Tumoricidal polypeptide(s) - in which DNA to code the polypeptide(s) is
 PT constructed from human-derived TNF-coding DNA and rabbit-derived
 PT TNF-coding DNA.
 PS Disclosure; Page 4; 11pp; Japanese.
 CC The sequence is derived from tumour necrosis factor DNA.
 CC See also R05173-76 and R05189.
 SQ Sequence 154 AA;
 Query Match 4.4%; Score 94; DB 2; Length 154;
 Best Local Similarity 29.8%; Pred. No. 3.18e+01;
 Matches 14; Conservative 14; Mismatches 18; Indels 1; Gaps 1;
 Db 31 anallangmkldnqlvvpadgyllysqvlfsgggcrsyvllthtv 77
 QY 53 THGLMA-GVTITATVILVSAGEAYLVYTDRLYSRDFNNVAAVYKV 98
 RESULT 10
 ID R03266 standard; protein; 154 AA.
 AC R03266;
 DT 18-JUL-1990 (first entry)
 DE Rabbit tumour necrosis factor polypeptide.
 KW Tumoricide; Tumour necrosis factor.
 OS Oryctolagus cuniculus.
 PN J02019398-A.
 PD 23-JAN-1990.
 PF 06-JUL-1988; 166913.
 PR 06-JUL-1988; JP-166913.
 PA (ASAH) Asahi Chemical Ind KK.

This Page Blank (uspto)


```
ACCESSIONS S70114
REFERENCE S70114
#authors Fulton, L.
#submission submitted to the EMBL Data Library, May 1996
#description The sequence of S. cerevisiae cosmid 9819.
#accession S70114
#molecule_type DNA
#residues 1-289 #label FUL
#cross-references EMBL:U51031
GENETICS
#map_position 4R
SUMMARY #length 289 #molecular-weight 33514 #checksum 4703
Query Match 6.8%; Score 144; DB 11; Length 289;
Best Local Similarity 24.8%; Pred. No. 1.76e-05;
Matches 58; Conservative 65; Mismatches 92; Indels 19; Gaps 18;
Db 19 ledvf-lllilmilnpyvyyqqfergfnditishpy-a-tter-vnnmlfvyfsf- 73
QY 9 LLDVLCLLVASLPFAILTLVNAFYKRGFYCGDDSIIRYPYRPTITHGLMAGVITATVIL 68
Db 74 vps-ltlllgsladrhli-fi-lytallgslawfststftfnkwngrlpdfl 129
QY 69 VSAGEAYLVTDRLYGRSDFNNVAAVYKVLGTLFGAAVSQSLTLDAKYMIGRLKPNFL 128
Db 130 droqpevglpd-tilftakd-vcttknherllldgfttp-sghssesfaglglyfwlwg 186
QY 129 AVCDDP-DWSRVNCSVVOLEKVCGRNPAD-VTEA-RLSFYSGHSSFGMYCMVFLAYVQA 185
Db 187 qltlesplmplkwmvafiplllgaal-ialstqdyrhhfvdvlgsmlygima 239
QY 186 RLCWKWARL-L-RPTVOFF-LVAFALYVGTRYSDYKHHWSDVLVLGLOALVA 236
RESULT 3
ENTRY D64233 #type complete
TITLE hypothetical protein MG302 - Mycoplasma genitalium (SGC3)
ORGANISM #formal_name Mycoplasma genitalium
DATE 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change
10-Nov-1995
ACCESSIONS D64233
REFERENCE A64200
#authors Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.; Sutton, G.; Kelley, J.M.; Pritchman, J.L.; Weidman, J.F.; Small, K.V.; Sandusky, M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.M.; Tomb, J.F.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.; Lucier, T.S.; Peterson, S.N.; Smith, H.O.; Hutchison III, C.A.; Venter, J.C.
#journal Science (1995) 270:397-403
#title The minimal gene complement of Mycoplasma genitalium.
#accession D64233
#status preliminary
#molecule_type DNA
#residues 1-317 #label TIGR
#cross-references GB:I43967; TIGR:MG302
#experimental_source strain G-37
#note neither nucleotide sequence nor conceptual translation is shown
GENETICS
#genetic_code SGC3
SUMMARY #length 317 #molecular-weight 36667 #checksum 8963
Query Match 5.3%; Score 112; DB 10; Length 317;
Best Local Similarity 24.5%; Pred. No. 1.44e-01;
Matches 25; Conservative 27; Mismatches 47; Indels 3; Gaps 3;
Db 167 islllavikllptkseiirikqatqrqfylnkcsfnlpfiktllfpvllstvkkte 226
QY 6 VFVLLDVLCLLVASLPFAILTLVNAFYKRGFYCGDDSIIRYPYRPTI-THGLMAGVTITA 64
Db 227 ttatfalcakgydinntnrthypikyn-llngvflvllfls 267
QY 65 TVTILVSAGEAY-LVYTDRLYSRSDFNNNVAAVYKVLGTLFG 105
RESULT 4
ENTRY SZBSSE #type complete
TITLE stage V sporulation protein E - Bacillus subtilis
ALTERNATE_NAMES spoVE protein
ORGANISM #formal_name Bacillus subtilis
DATE 28-Dec-1987 #sequence_revision 06-Jan-1995 #text_change
06-Jan-1995
ACCESSIONS S10243; A29756; E47691; S22211
REFERENCE S10243
#authors Sato, T.; Theeragool, G.; Yamamoto, T.; Okamoto, M.; Kobayashi, Y.
#journal Nucleic Acids Res. (1990) 18:4021
#title Revised nucleotide sequence of the sporulation gene spoVE from Bacillus subtilis.
#cross-references MUID:90326557
#accession S10243
#status translation not shown
#molecule_type DNA
#residues 1-366 #label SAT
#cross-references EMBL:X51419
REFERENCE A29756
#authors Bugaichuk, U.D.; Piggot, P.J.
#journal J. Gen. Microbiol. (1986) 132:1883-1890
#title Nucleotide sequence of the Bacillus subtilis developmental gene spoVE.
#cross-references MUID:87085411
#accession A29756
#molecule_type DNA
#residues A47691
#authors Daniel, R.A.; Errington, J.
#journal J. Gen. Microbiol. (1993) 139:361-370
#title DNA sequence of the mreE-murD region of Bacillus subtilis 168.
#cross-references MUID:93171879
#contents 168
#accession E47691
#molecule_type DNA
#residues 1-40 #label DAN
#cross-references NCBI:N125659; NCBI:P125665
#note sequence extracted from NCBI backbone
REFERENCE S22211
#authors Henriques, D.O.; de Lencastre, H.; Piggot, P.J.
#submission submitted to the EMBL Data Library, January 1992
#accession S22211
#molecule_type DNA
#residues 343-366 #label HEN
#cross-references EMBL:X64259
GENETICS
#gene spoVE
#map_position 133 (degrees)
#start_codon TTG
CLASSIFICATION #superfamily rod shape-determining protein
KEYWORDS sporulation
SUMMARY #length 366 #molecular-weight 40132 #checksum 1239
Query Match 5.1%; Score 107; DB 4; Length 366;
Best Local Similarity 22.6%; Pred. No. 5.27e-01;
Matches 19; Conservative 25; Mismatches 35; Indels 5; Gaps 5;
Db 53 agigviam-ffimvdywt-wrtwskllm-vicffllvllpvgvmvngerswlgvga 109
QY 169 SSFGMYCMVFLAYVQARLCKWK-WARLLRPTVOFFLVAFALYVGTRYSDYKHHWSDVLV 227
Db 110 fsiqsefneklamiaflakflek 133
QY 228 GLLOQALVAALT-VCIYISDFFKAR 250
```



```

Query Match      4.9%; Score 103; DB 13; Length 477;
Best Local Similarity 26.5%; Pred. No. 1.45e+00;
Matches          22; Conservative 30; Mismatches 25; Indels 6; Gaps 6;

Db    276 llgqdvtaismftslsvaatgflplssaiysrllsihelthvpskiltglf-laip1 334
QY    56 LMAQ-VTITATV-ILVS-AGEAYLVYTDRLYSRS-DFNIV-AAVKYVLGTFLFGAANSQ 110
Db    335 avgvlisklpkfsgllqvvpk 357
QY    111 SLTDLAKYMIGRLKPNFLAVCDP 133

RESULT 11
ENTRY   S69990 #type complete
TITLE   YKR106w homolog YCL070/73c - Yeast (Saccharomyces cerevisiae)
ORGANISM #formal_name Saccharomyces cerevisiae
DATE     05-Dec-1996 #sequence_revision 05-Dec-1996 #text_change
ACCESSIONS S69990
REFERENCE   S69990
#authors   Gromadka, R.; Gora, M.; Zielenkiewicz, U.; Slonimski, P.P.;
            Rytko, J.
#journal   yeast (1996) 12:593-591
#title     Subtelomeric duplications in Saccharomyces cerevisiae
            chromosomes III and XI: topology, arrangements, corrections
            of sequence and strain-specific polymorphism.
#accession S69990
#status     preliminary
#residues   1-615 ##label GRO
##cross-references EMBL:X59720
SUMMARY    #length 615 #molecular-weight 68897 #checksum 9357

Query Match      4.9%; Score 103; DB 16; Length 615;
Best Local Similarity 23.5%; Pred. No. 1.45e+00;
Matches          23; Conservative 34; Mismatches 33; Indels 8; Gaps 7;

Db    47 eimsaqfdislkykil-listafvcfgisldytlrstytgyatnsysehsllstqvina 105
QY    11 DVLCLLVASLPFALLTLVNAPYKRGF-YCGDDSIIR--YP-YRPDTIT-HGLMAGVTITAT 65
Db    106 vsvagsqvvyrsldh-fgrlrif-lvatifylmgtii 141
QY    66 VILVSAGEALVTVTDRLYSRSDFNNTVAAYKVLTGTF 103

RESULT 12
ENTRY   S40624 #type complete
TITLE   hypothetical protein YCL070C homolog YKR106w - yeast
            [Saccharomyces cerevisiae]
ORGANISM #formal_name Saccharomyces cerevisiae
DATE     03-May-1994 #sequence_revision 03-May-1994 #text_change
ACCESSIONS S40624
REFERENCE   S40624
#authors   Gaillon, L.; Dujon, B.
#submission submitted to the Protein Sequence Database, March 1994
#accession S40624
#molecule_type DNA
#residues   1-615 ##label GAI
##cross-references EMBL:Z28202
##experimental_source strain 5288C
GENETICS
#map_position 11R
SUMMARY    #length 615 #molecular-weight 68953 #checksum 2974

Query Match      4.9%; Score 103; DB 11; Length 615;
Best Local Similarity 23.5%; Pred. No. 1.45e+00;
Matches          23; Conservative 34; Mismatches 33; Indels 8; Gaps 7;

Db    47 eimsaqfdislkykil-listafvcfgisldytlrstytgyatnsysehsllstqvina 105
QY    11 DVLCLLVASLPFALLTLVNAPYKRGF-YCGDDSIIR--YP-YRPDTIT-HGLMAGVTITAT 65

Query Match      4.9%; Score 103; DB 13; Length 367;
Best Local Similarity 18.8%; Pred. No. 1.45e+00;
Matches          15; Conservative 32; Mismatches 30; Indels 3; Gaps 3;

Db    155 vsnallsvaisvilitvstyls-pqvilvk-dyklnficqvtgmfsfyigvvrn 212
QY    52 ITHGLMAGVTITATVLSAGEAYLVYTDRLYSRSDFN-NYAAYKVLTGTFPGAANSQ 110
Db    213 qlynlfnfyvfilltvlily 232
QY    111 SLTDLAKYMIGRLKPNFLAV 130

RESULT 9
ENTRY   S19402 #type complete
TITLE   hypothetical protein YCL070c - yeast (Saccharomyces cerevisiae)
ORGANISM #formal_name Saccharomyces cerevisiae
DATE     31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
ACCESSIONS S19402
REFERENCE   S19402
#authors   van der Aart, Q.J.M.; Steensma, H.Y.
#submission submitted to the Protein Sequence Database, March 1992
#accession S19402
#molecule_type DNA
#residues   1-373 ##label AAR
##cross-references EMBL:X59720
GENETICS
#map_position 3L
SUMMARY    #length 373 #molecular-weight 42503 #checksum 4433

Query Match      4.9%; Score 103; DB 11; Length 373;
Best Local Similarity 23.5%; Pred. No. 1.45e+00;
Matches          23; Conservative 34; Mismatches 33; Indels 8; Gaps 7;

Db    47 eimsaqfdislkykil-listafvcfgisldytlrstytgyatnsysehsllstqvina 105
QY    11 DVLCLLVASLPFALLTLVNAPYKRGF-YCGDDSIIR--YP-YRPDTIT-HGLMAGVTITAT 65
Db    106 vsvagsqvvyrsldh-fgrlrif-lvatifylmgtii 141
QY    66 VILVSAGEALVTVTDRLYSRSDFNNTVAAYKVLTGTF 103

RESULT 10
ENTRY   S01696 #type complete
TITLE   gene P3 protein - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE     30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change
ACCESSIONS S01696
REFERENCE   S01696
#authors   Alcalay, M.; Toniolo, D.
#journal   Nucleic Acids Res. (1988) 16:9527-9543
#title     CpG islands of the X chromosome are gene associated.
#cross-references MUID:89041548
#accession S01696
#molecule_type DNA
#residues   1-477 ##label ALC
##cross-references EMBL:X12458
SUMMARY    #length 477 #molecular-weight 50332 #checksum 7015

```



```

Db 106 vsvsgqvvyrlsdh-fgrlrif-lvlatifimgtll 141
Qy 66 VILVSAGEAYLVYTDRLYSRSDFNYYAAVKVGLGTFL 103

RESULT 13
ENTRY S12686 #type complete
TITLE coat protein - belladonna mottle virus
ALTERNATE_NAMES virion protein
ORGANISM #formal_name belladonna mottle virus
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
ACCESSIONS S12686
REFERENCE S12686
#authors Ding, S.W.; Howe, J.; Mackenzie, A.; Skotnicki, M.; Gibbs, A.
#journal Nucleic Acids Res. (1990) 18:6138
#title Nucleotide sequence of the virion protein gene of belladonna
mottle tymovirus.
#cross-references MUID:91045090
#accession S12686
#molecule_type DNA
#residues 1-188 #label DIN
#cross-references EMBL:X54529
#note the authors translated the codon ACC for residue 64 as U
and ACA for residue 167 as U
CLASSIFICATION #superfamily turnip yellow mosaic virus coat protein
SUMMARY #length 188 #molecular-weight 19891 #checksum 5832

Query Match 4.8%; Score 101; DB 8; Length 188;
Best Local Similarity 24.1%; Pred. No. 2.38e+00;
Matches 20; Conservative 28; Mismatches 30; Indels 5; Gaps 5;

Db 37 qfeatsvgyetlaq-v-nlssdsiakltgyrakvvelftitptgra-idcpvtvd 93
Qy 87 DENNYVAAYKVLGFLFGAAVQSILDLAK-YMIGRLKPNFLAVCDPDWSEVNCVYVQ 145

Db 94 vawv-panstappskillslygq 115
Qy 146 LEKVCGRNPADVTEARLSFYSGH 168

RESULT 14
ENTRY S24611 #type complete
TITLE latent membrane protein - human herpesvirus 4
ORGANISM #formal_name human herpesvirus 4, Epstein-Barr virus
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
ACCESSIONS S24611
REFERENCE S24611
#authors Chang, Y.S.
#submission submitted to the EMBL Data Library, June 1992
#accession S24611
#status preliminary
#molecule_type DNA
#residues 1-381 #label CHA
#cross-references EMBL:X66863

GENETICS
#introns 90/1; 119/1
CLASSIFICATION #superfamily Epstein-Barr virus latent membrane protein
SUMMARY #length 381 #molecular-weight 41372 #checksum 3544

Query Match 4.7%; Score 100; DB 8; Length 381;
Best Local Similarity 25.7%; Pred. No. 3.04e+00;
Matches 18; Conservative 21; Mismatches 27; Indels 4; Gaps 4;

Db 104 alylgivlif-ifgcilvlgwlyleilwrlgatlwgllafil-afflaillilialyl 161
Qy 163 SFYSGHSSFGMY-CMVFLALYVQARLCWKARLLRPTVQFFLVAFALYGVYTRVSDY-KH 220

Db 162 nwtllvdl 171
Qy 221 HWSDLVGLL 230

```

```

RESULT 15
ENTRY LABECA #type complete
TITLE latent membrane protein - human herpesvirus 4 (strain CAO)
ORGANISM #formal_name human herpesvirus 4, Epstein-Barr virus
DATE 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change
ACCESSIONS JQ1434
REFERENCE JQ1434
#authors Hu, L.F.; Zabarovsky, E.R.; Chen, F.; Cao, S.L.; Ernberg, I.;
Klein, G.; Winberg, G.
#journal J. Gen. Virol. (1991) 72:2399-2409
#title Isolation and sequencing of the Epstein-Barr virus BNLF-1
gene (LMPL) from a Chinese nasopharyngeal carcinoma.
#cross-references MUID:92013956
#accession JQ1434
#molecule_type DNA
#residues 1-404 #label HUL
#note the authors translated the codon AAA for residue 358 as
Ala
GENETICS
#gene LMPL; BNLF-1
#introns 90/1; 119/1
CLASSIFICATION #superfamily Epstein-Barr virus latent membrane protein
KEYWORDS transmembrane protein
FEATURE
25-45 #domain transmembrane #status predicted #label TM1\
51-71 #domain transmembrane #status predicted #label TM2\
76-97 #domain transmembrane #status predicted #label TM3\
104-124 #domain transmembrane #status predicted #label TM4\
140-160 #domain transmembrane #status predicted #label TM5\
166-186 #domain transmembrane #status predicted #label TM6\
256-332 #region 11-residue repeats (P-D-N-T-D-D-N-G-P-Q-D)
SUMMARY #length 404 #molecular-weight 43769 #checksum 7535

Query Match 4.6%; Score 97; DB 4; Length 404;
Best Local Similarity 29.2%; Pred. No. 6.28e+00;
Matches 21; Conservative 18; Mismatches 26; Indels 7; Gaps 7;

Db 104 alylgivlif-ifgcilvlgwlyleilwrlgatlwgllafil-afflaillilialyl 160
Qy 163 SFYSGHSSFGMY-CM-VF-LALYVQARLCWKARLLRPTVQFFLVAFALYGVYTRVSDY- 218

Db 161 qgnwttllvdl 172
Qy 219 KHRWSDVLVGLL 230

Search completed: Tue Nov 4 10:32:01 1997
Job time : 77 secs.

```

This Page Blank (uspto)

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

```

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Nov 4 10:29:23 1997; MasPar time 7.58 Seconds
Tabular output not generated. 772.348 Million cell updates/sec

```

```
>US-08-842-827-8
Title:
Description: (1-276) from US08842827.pap
Perfect Score: 2118
Sequence: 1 MORRVFVLLDVLICLVASL.....KEEELERKPSISLTLTILGRG 276
```

Scoring table: PAM 150
Gap 11

Searched: 59021 seqs, 21210388 residues

Post-processing: Minimum Match 08
Listing first 45 summaries

```
Database:      swiss-prot34
               1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
               8:part8 9:part9 10:part10 11:part11
```

Statistics: Mean 48.002; Variance 91.621; scale 0.524

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query		Score	Match	Length	DB	ID	Description	Pred. No.
	Score	Match							
1	461	21.8	341	11	YSX3_CAGEEL	HYPOTHETICAL 39.0 KD	2.95e-67		
2	112	5.3	317	10	YSO2_MYGE	HYPOTHETICAL PROTEIN	1.65e-02		
3	107	5.1	366	9	SPSE_BACSU	STAGE V SPOULRULATION P	7.74e-02		
4	108	5.1	596	11	YB42_YEAST	HYPOTHETICAL AMINO-AC	5.70e-02		
5	105	5.0	386	6	LMP1_EBVV	LATENT MEMBRANE PROTE	1.42e-01		
6	103	4.9	332	2	CCPA_BACME	GLUCOSE-RESISTANCE AM	2.57e-01		
7	103	4.9	477	7	P3_HUMAN	P3 PROTEIN.	2.57e-01		
8	103	4.9	615	11	YCHO_YEAST	HYPOTHETICAL 68.9 KD	2.57e-01		
9	103	4.9	615	11	YK86_YEAST	HYPOTHETICAL 69.0 KD	2.57e-01		
10	99	4.7	309	11	YOGH_BACSU	PROBABLE ABC TRANSPOR	8.29e-01		
11	98	4.6	261	11	YV23_MYCLE	HYPOTHETICAL 27.9 KD	1.10e+00		
12	97	4.6	404	6	LMP1_EBVV	LATENT MEMBRANE PROTE	1.47e+00		
13	97	4.6	616	11	YGO2_YEAST	HYPOTHETICAL 72.6 KD	1.47e+00		
14	97	4.6	692	6	NGTL_NAAT	SODIUM- AND CHLORIDE	1.47e+00		
15	95	4.5	2019	2	CIN5_RAT	SODIUM CHANNEL PROTEI	2.57e+00		
16	94	4.4	233	9	TNFA_BOVIN	TUMOR NECROSIS FACTOR	3.39e+00		
17	94	4.4	235	9	INF4_RABIT	TUMOR NECROSIS FACTOR	3.39e+00		
18	94	4.4	334	2	CCPA_BACSU	GLUCOSE-RESISTANCE AM	3.39e+00		
19	94	4.4	386	6	LMP1_EBV	LATENT MEMBRANE PROTE	3.39e+00		
20	94	4.4	490	1	ACM4_CHICK	MUSCARINIC ACETYLCHOL	3.39e+00		
21	94	4.4	555	6	NRM1_CHICK	NATURAL RESISTANCE-AS	3.39e+00		
22	93	4.4	633	6	NGM1_RAT	SODIUM- AND CHLORIDE-	4.47e+00		

QY	163	SPYSGHSGMY-CHVFLLAL--YVQARLCWKWARLLRPTQVFFLVFALYVGVTRVSDY - 218
Db	161	qgqwwtllvdl1 172 : : :
QY	219	KHNSDVLVGLL 230 : : :
RESULT	6	
ID	CCPA_BACME	STANDARD; PRT; 332 AA.
AC	P46828;	
DT	01-NOV-1995	(REL. 32, CREATED)
DT	01-NOV-1995	(REL. 32, LAST SEQUENCE UPDATE)
DT	01-OCT-1996	(REL. 34, LAST ANNOTATION UPDATE)
GN	GLUCOSE-RESISTANCE AMYLASE REGULATOR (CATABOLITE CONTROL PROTEIN).	
CC	CCPA.	
OC	BACILLUS MEGATERIUM.	
OS	PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE; 94259294.	
RL	HUECK C., KRAUS A., HILLEN W.;	
RL	GENE 143:147-148(1994).	
CC	-!- FUNCTION: TRANSCRIPTIONAL REGULATOR INVOLVED IN BOTH THE	
CC	REPRESSION OF CARBOHYDRATE UTILIZATION GENES SUCH AS THE ALPHA-	
CC	AMYLASE (AMYE) AND THE ACETYL-COENZYME A SYNTHETASE (ACSA); AND IN	
CC	THE POSITIVE REGULATION OF GENES INVOLVED IN EXCRETION OF EXCESS	
CC	CARBON SUCH AS THE ACETATE KINASE (ACKA). MAY PLAY A ROLE IN	
CC	REGULATION OF CITRATE SYNTHASE GENE EXPRESSION.	
CC	-!- SIMILARITY: BELONGS TO THE LACI FAMILY OF TRANSCRIPTIONAL	
CC	REGULATORS.	
CC	EMB1; L26052; G415664; "	
DR	PROSITE; PS00356; HTH_LACI_FAMILY.	
DR	TRANSCRIPTION REGULATION; REPRESSOR; ACTIVATOR: DNA-BINDING.	
FW	DNA_BIND 5 24 H-T-H MOTIF (POTENTIAL).	
QY	SEQUENCE 332 AA; 36644 MW; A174D030 CRC32;	
Query Match	4.9%;	Score 103; DB 2; Length 332;
Best Local Similarity	26.2%;	Pred. No. 2.57e-01;
Matches	22; Conservative	28; Mismatches 25; Indels 9; Gaps
Db	39	ettierlgypnavarglaskkttvgvipdlsnif--yae-l-arg-led-iatmyk-y 91 : : : : : : : : : : : : : : : : : :
QY	41	DSI-RYPYRPDTHIGLMAGVITATVILVSAGEAYLVYTRDLYSRDFNNYAAVYKVL 99 : : : : : : : : : : : : : :
Db	92	niilnsdgnqdkelhlnnmlqk 115 : : : : : : : : : : : : :
QY	100	GTFLFGAAVSQSLT-DLAKYMIGR 122 : : : : : : : : : : : : :
RESULT	7	
ID	P3_HUMAN	STANDARD; PRT; 477 AA.
AC	P09131;	
DT	01-MAR-1989	(REL. 10, CREATED)
DT	01-MAR-1989	(REL. 10, LAST SEQUENCE UPDATE)
DT	01-OCT-1996	(REL. 34, LAST ANNOTATION UPDATE)
DE	P3 PROTEIN.	
GN	P3.	
OS	HOMO SAPIENS (HUMAN).	
OS	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;	
OC	EUTHERIA; PRIMATES.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE; 89041548.	
RA	ALCALAY M., TONIOLO D.;	
RL	NUCLEIC ACIDS RES. 16:9527-9556(1988).	
RL	[2]	
RP	SEQUENCE FROM N.A.	
RP	CHEN E.Y., ZOLLO M., MAZZARELLA R.A., CICCOCICOLA A., CHEN C.N.,	
RA	ZUO L., HEINER C., BURROGH F.W., RIETTO M., SCHLESSINGER D.,	
RA	D'URSO M.;	
RL	SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.	
CC	-!- FUNCTION: THE UBIQUITOUS EXPRESSION AND THE CONSERVATION OF THE	
CC	P3 SEQUENCE IN DISTANT ANIMAL SPECIES SUGGEST THAT THE P3 GENE	

CC CODES FOR A PROTEIN WITH HOUSEKEEPING FUNCTIONS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: TO P3 PROTEIN OF ANIMALS AND YEASTS.
CC -1- SIMILARITY: BELONGS TO THE SODIUM-BILE ACID SYMPORTER FAMILY (SBF).

DR EMBL: X12458; G35188; -.
DR EMBL: L44140; G1203976; -.
DR PIR: S01696; S01696.
DR MIM: 312090; .

KW TRANSMEMBRANE; TRANSPORT; SYMPORT.
SQ SEQUENCE 477 AA; 50332 MW; 653D4BA7 CRC32;

Query Match 4.9%; Score 103; DB 7; Length 477;

Best Local Similarity 26.5%; Pred. No. 2.57e-01;

Matches 22; Conservative 30; Mismatches 25; Indels 6; Gaps 6;

Db 276 llogdvtaismfistvaatgfpilssayrlslshethlpvdkilgtllf-taapl 334

QY 56 LMAQ-VTITATV-ILVS-AGEAYLVYTDRLYSRSDFNNTV-AAVYKVLGTFLFGAASQ 110

Db 335 avgvliksklpkfsqlllqvvp 357

QY 111 SLTDLAKYMGRLKPNFLAVCDP 133

RESULT 8

ID YCHO YEAST STANDARD; PRT; 615 AA.

AC P25596; P25597; P25599;

DT 01-MAY-1992 (REL. 22, CREATED)

DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)

DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)

DE HYPOTHETICAL 68.9 KD PROTEIN IN HML 5'REGION.

GN YCLO70C/71C/73C.

OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).

OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.

RN [1]

RP SEQUENCE FROM N.A.

RA VAN DER AART Q.J.M., STEENSMA H.Y.;

RL SUBMITTED (MAR-1992) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [2]

RP REVISIONS.

RA GROMADKA R.;

RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC -1- SIMILARITY: BELONGS TO THE YCL70C/YHL047C/YKR106W FAMILY.

DR EMBL: X59720; E264591; .

DR PIR: S19402; S19402.

DR PIR: S19403; S19403.

DR PIR: S19405; S19405.

KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.

FT TRANSMEM 59 79 POTENTIAL.

FT TRANSMEM 121 141 POTENTIAL.

FT TRANSMEM 153 173 POTENTIAL.

FT TRANSMEM 187 207 POTENTIAL.

FT TRANSMEM 217 237 POTENTIAL.

FT TRANSMEM 276 296 POTENTIAL.

FT TRANSMEM 308 328 POTENTIAL.

FT TRANSMEM 344 364 POTENTIAL.

FT TRANSMEM 384 404 POTENTIAL.

FT TRANSMEM 409 429 POTENTIAL.

FT TRANSMEM 441 461 POTENTIAL.

FT TRANSMEM 472 492 POTENTIAL.

FT TRANSMEM 548 568 POTENTIAL.

SQ SEQUENCE 615 AA; 68898 MW; 0BBD5B37 CRC32;

Query Match 4.9%; Score 103; DB 11; Length 615;

Best Local Similarity 23.5%; Pred. No. 2.57e-01;

Matches 23; Conservative 34; Mismatches 33; Indels 8; Gaps 7;

Db 47 eimsaqfslkylil-listafvcgfgisldyltrstgyatnsysehslstvgvina 105

QY 11 DVLCLLVASLPFAILLVNAPYKRGF-YCGDDSIIR--YP-YRPDTIT-HGLMAGVTITAT 65

Db 106 vsvsgsqvvyrsldsh-fgrlrlf-lvatlfymgtll 141

QY 66 VILVSAGEAYLVYTDRLYSRSDFNNTV-AAVYKVLGTFL 103

RESULT 9

ID YK86 YEAST STANDARD; PRT; 615 AA.

AC P36173;

DT 01-JUN-1994 (REL. 29, CREATED)

DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)

DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)

DE HYPOTHETICAL 69.0 KD PROTEIN IN SIR1 3'REGION.

GN YKR106W.

OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).

OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.

RN [1]

RP SEQUENCE FROM N.A.

RA GAILLON L., DUJON B.;

RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC -1- SIMILARITY: BELONGS TO THE YCL70C/YHL047C/YKR106W FAMILY.

DR EMBL: Z28202; G486617; .

DR PIR: S40624; S40624.

KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.

FT TRANSMEM 59 79 POTENTIAL.

FT TRANSMEM 121 141 POTENTIAL.

FT TRANSMEM 153 173 POTENTIAL.

FT TRANSMEM 217 237 POTENTIAL.

FT TRANSMEM 276 296 POTENTIAL.

FT TRANSMEM 308 328 POTENTIAL.

FT TRANSMEM 344 364 POTENTIAL.

FT TRANSMEM 384 404 POTENTIAL.

FT TRANSMEM 408 428 POTENTIAL.

FT TRANSMEM 441 461 POTENTIAL.

FT TRANSMEM 472 492 POTENTIAL.

FT TRANSMEM 549 569 POTENTIAL.

SQ SEQUENCE 615 AA; 68954 MW; A8423CFB CRC32;

Query Match 4.9%; Score 103; DB 11; Length 615;

Best Local Similarity 23.5%; Pred. No. 2.57e-01;

Matches 23; Conservative 34; Mismatches 33; Indels 8; Gaps 7;

Db 47 eimsaqfslkylil-listafvcgfgisldyltrstgyatnsysehslstvgvina 105

QY 11 DVLCLLVASLPFAILLVNAPYKRGF-YCGDDSIIR--YP-YRPDTIT-HGLMAGVTITAT 65

Db 106 vsvsgsqvvyrsldsh-fgrlrlf-lvatlfymgtll 141

QY 66 VILVSAGEAYLVYTDRLYSRSDFNNTV-AAVYKVLGTFL 103

RESULT 10

ID YQGH_BACSU STANDARD; PRT; 309 AA.

AC P46339;

DT 01-NOV-1995 (REL. 32, CREATED)

DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)

DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)

DE PROBABLE ABC TRANSPORTER PERMEASE PROTEIN IN SODA-COMGA INTERGENIC REGION (ORF72).

GN YQGH.

OS BACILLUS SUBTILIS.

OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-168 / JH642;

RA TAKEMARU K., MIZUNO M., SATO T., TAKEUCHI M., KOBAYASHI Y.;

RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-168 / JH642;

RA KOBAYASHI Y., MIZUNO M., MASUDA S., TAKEMARU K., HOSONO S.,

RA SATO T., TAKEUCHI M.;

RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

FT TRANSMEM 424 444 POTENTIAL.
 FT CONFLICT 339 S -> A (IN REF. 2).
 SQ SEQUENCE 616 AA: 72565 MW; CBE29FT0 CRC32;

Query Match 4.6%; Score 97; DB 11: Length 616;
 Best Local Similarity 27.5%; Pred. No. 1.47e+00;
 Matches 22; Conservative 22; Mismatches 30; Indels 6; Gaps 6;

Db 87 illhyal-lllslgslldlllllpkylsvqvaedklrlpfdvtrsfeyggyviyapklvm 145
 QY 8 VLLD-VCLLVASLPAILLVNAPYKRGYCGDSTRYPYRPD-RITH-GLMAGVTITA 64

Db 146 av-laslggyivrfvqkly 164
 QY 65 TVILVSAGEAYLV-YTDRLY 83

RESULT 14
 ID NTGL-HUMAN STANDARD; PRT: 692 AA.
 AC P48067;
 DT 01-FEB-1996 (REL. 33, CREATED)
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE SODIUM- AND CHLORIDE-DEPENDENT GLYCINE TRANSPORTER 1 (GLYT-1).
 GN SLC6A9.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; PRIMATES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RX MEDLINE; 94239375.
 RA KIM K.-M., KINGSMORE S.F., HAN H., YANG-FENG T.L., GODINOT N.,
 RA SELDIN M.F., CARON M.G., GIROS B.;
 RL MOL. PHARMACOL. 45:608-617(1994).
 CC -1- FUNCTION: TERMINATES THE ACTION OF GLYCINE BY ITS HIGH AFFINITY
 CC SODIUM-DEPENDENT REUPTAKE INTO PRESYNAPTIC TERMINALS.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: ALL GLYT-1 SUBTYPES CAN BE FOUND IN BRAIN,
 CC KIDNEY, PANCREAS, LUNG, PLACENTA, LIVER. BUT GLYT-1C SUBTYPE IS
 CC ONLY FOUND IN THE BRAIN
 CC -1- ALTERNATIVE PRODUCTS: THREE ISOFORMS (GLYT-1A, GLYT-1B, AND GLYT-
 CC 1C) ARE PRODUCED BY ALTERNATIVE SPLICING. FORM GLYT-1C IS SHOWN
 CC HERE.
 CC -1- SIMILARITY: BELONGS TO THE SODIUM:NEUROTRANSMITTER SYMPORTER
 CC FAMILY (SNF).
 CC EMBL; S70609; G546769; -;
 DR EMBL; S70612; G546771; -;
 DR MIM; 601019; -;
 DR KIM; 601019; -;
 KW NEUROTRANSMITTER TRANSPORT; TRANSPORT; TRANSMEMBRANE; GLYCOPROTEIN;
 KW SYMPORT; AMINO-ACID TRANSPORT; ALTERNATIVE SPLICING.
 FT DOMAIN 1 94
 FT TRANSMEM 95 115
 FT TRANSMEM 123 142
 FT TRANSMEM 166 186
 FT DOMAIN 187 273
 FT TRANSMEM 274 292
 FT TRANSMEM 301 318
 FT TRANSMEM 354 371
 FT TRANSMEM 383 404
 FT TRANSMEM 437 456
 FT TRANSMEM 485 503
 FT TRANSMEM 519 539
 FT TRANSMEM 560 579
 FT TRANSMEM 598 616
 FT DOMAIN 618 692
 FT CARBOHYD 223 226
 FT CARBOHYD 226 226
 FT CARBOHYD 236 236
 FT CARBOHYD 242 242
 FT VARSPLIC 1 69
 MAAHGPVAPSSPEQVTLTPVQRSLPFPFGATPSTSLAE
 SVLKVWHGATNSGLFQMLAQHSLAMAAQ -> MVGKGAKGM
 L (IN FORM 1A).

FT VARSPLIC 16 69 MISSING (IN FORM 1B).
 SQ SEQUENCE 692 AA: 76823 MW; 5A5184C7 CRC32;

Query Match 4.6%; Score 97; DB 6: Length 692;
 Best Local Similarity 28.3%; Pred. No. 1.47e+00;
 Matches 15; Conservative 17; Mismatches 14; Indels 7; Gaps 7;

Db 562 wrfvspalifilvftv-ldqypitynhyqpgwa-vaigflm-alsvvlcl 610
 QY 191 WARLLRPTVQFLVAFALVGVTRVS-D-YKHH-WSDVLVGLLQALVAALIV 240

RESULT 15
 ID CIN5-RAT STANDARD; PRT: 2019 AA.
 AC P15389;
 DT 01-APR-1990 (REL. 14, CREATED)
 DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE SODIUM CHANNEL PROTEIN, CARDIAC MUSCLE ALPHA-SUBUNIT.
 GN SCN5A.
 OS RATTUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; RODENTIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HEART;
 RX MEDLINE; 90046760.
 RA ROGART R.B., CRIBBS L.L., MUGLIA L.K., KEPHART D.D., KAISER M.W.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 86:8170-8174(1989).
 CC -1- FUNCTION: THIS PROTEIN MEDIATES THE VOLTAGE-DEPENDENT SODIUM ION
 CC PERMEABILITY OF EXCITABLE MEMBRANES. ASSUMING OPENED OR CLOSED
 CC CONFORMATION IN RESPONSE TO THE VOLTAGE DIFFERENCE ACROSS THE
 CC MEMBRANE, THE PROTEIN FORMS A SODIUM-SELECTIVE CHANNEL THROUGH
 CC WHICH NA+ IONS MAY PASS IN ACCORDANCE WITH THEIR ELECTROCHEMICAL
 CC GRADIENT. IT IS A TETRODOTOXIN-RESISTANT NA+ CHANNEL ISOFORM.
 CC -1- NA+ CHANNELS IN MAMMALIAN CARDIAC MEMBRANE HAVE FUNCTIONAL
 CC PROPERTIES QUITE DISTINCT FROM NA+ CHANNELS IN NERVE AND SKELETAL
 CC MUSCLE.
 CC -1- DOMAIN: THE SEQUENCE CONTAINS 4 INTERNAL REPEATS, EACH WITH 5
 CC HYDROPHOBIC SEGMENTS (S1,S2,S3,S5,S6) AND ONE POSITIVELY CHARGED
 CC SEGMENT (S4). SEGMENTS S4 ARE PROBABLY THE VOLTAGE-SENSORS AND ARE
 CC CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT
 CC EVERY THIRD POSITION.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: TO OTHER SODIUM CHANNEL PROTEINS.
 CC EMBL; M27902; G206858; -;
 DR PIR; A33996; A33996.
 KW IONIC CHANNEL; TRANSMEMBRANE; ION TRANSPORT; VOLTAGE-GATED CHANNEL;
 KW GLYCOPROTEIN; DUPLICATION; MULTIGENE FAMILY; PHOSPHORYLATION.
 FT TRANSMEM 128 151
 FT TRANSMEM 160 179
 FT TRANSMEM 193 211
 FT TRANSMEM 218 237
 FT TRANSMEM 254 277
 FT TRANSMEM 391 416
 FT TRANSMEM 713 737
 FT TRANSMEM 749 772
 FT TRANSMEM 781 800
 FT TRANSMEM 807 826
 FT TRANSMEM 843 863
 FT TRANSMEM 917 942
 FT TRANSMEM 1203 1226
 FT TRANSMEM 1240 1265
 FT TRANSMEM 1272 1293
 FT TRANSMEM 1298 1319
 FT TRANSMEM 1339 1361
 FT TRANSMEM 1446 1472
 FT TRANSMEM 1526 1549
 FT TRANSMEM 1561 1584
 FT TRANSMEM 1591 1614
 FT TRANSMEM 1635 1646
 FT TRANSMEM 1662 1684
 FT TRANSMEM 1750 1774

Search completed: Tue Nov 4 10:30:26 1997
Job time : 53 secs.

This Page Blank (uspto)